Système de télé-alarme médical : application à la détection des chutes de la personne âgée

Philippe Katz

THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE
sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de
DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE
Mention :Sciences et Technologies de l'Information et de la Communication - Spécialité Traitement du Signal et des Images
École Doctorale SICMA

présentée par
Philippe Katz
Préparée dans l'équipe VISION
ISEN Brest

Système de télé-alarme médical : application à la détection des chutes de la personne âgée

Thèse soutenue le 16 décembre 2014
devant le jury composé de :

Badr-Eddine BENKELFAT
Professeur, Telecom SudParis / *Rapporteur*

Denis HAMAD
Professeur, LISIC - ULCO Calais / *Rapporteur*

Ayman ALFALOU
Professeur, ISEN Brest / *Directeur de la thèse*

Michael ARON
Enseignant Chercheur, ISEN Brest / *Encadrant de la thèse*

Christian BROSSEAU
Professeur, Université de Bretagne Occidentale / *Président du jury*

Remerciements

Ce travail n'aurait pu arriver à son terme sans le support et la contribution morale, intellectuelle et matérielle d'un grand nombre de personnes et je les en remercie.

Plus particulièrement, je remercie M. Christian BROSSEAU, Professeur à l'Université de Bretagne Occidentale, pour avoir assumé la présidence du jury, ainsi que MM. Badr-Eddine BENKELFAT et Denis HAMAD, rapporteurs de ce manuscrit pour leurs conseils et remarques pertinents.

Mes remerciements vont à MM. Ayman ALFALOU et Michael ARON, respectivement directeur et encadrant de cette thèse, pour m'avoir encadré, conseillé et surtout encouragé tout au long de ces trois années.

Également je remercie le personnel de l'ISEN pour la formation scientifique qu'ils m'ont apportée pendant les années précédant cette thèse ainsi que pour les nombreuses discussions autour de la machine à café, tout particulièrement Caroline JEGAT, Alain "Tahiti Bob" LE BELLU, Nathalie ROUSSELET, Alain LOUSSERT, Yann RIOU et Pascal POUCHARD. Merci aussi à Dominique MARATRAY pour l'aide apportée dans la correction des articles.

Ma gratitude va en outre aux membres de l'équipe Vision, à Maher, Isabelle, Thibault, Marwa, Yousri pour les nombreuses discussions et remarques.

Je tiens à remercier mes compagnons d'infortune, les doctorants du laboratoire, pour m'avoir supporté pendant ce temps. Merci à Djamel, mon "co-détenu", pour les échanges de balles de tennis et l'invention du golf de bureau, à Angel pour les boulets de canon et les nombreuses discussions sur la société et enfin à Mohammed avec qui j'ai partagé les derniers jours d'un condamné avant la soutenance.

Un grand merci va à mes amis pour leur soutien durant ces trois années difficiles : Jérémy, Briac, Nicolas, Lucile, Quentin, Clémence, William, Catherine, Marine... Merci pour les cartes postales paradisiaques pendant que mon unique source lumineuse se résumait à un écran ! ! Je tiens à remercier également Laure sans qui je n'aurai certainement jamais réalisé ce travail, bien que nos chemins se soient aujourd'hui séparés.

Finalement, je remercie ma famille : mes parents, sans qui l'auteur de cette ligne n'aurait jamais vu le jour, mes grand-parents sans qui ce sont les auteurs de l'auteur qui ne seraient pas là (on peut remonter loin comme ça), mon frère, mes sœurs, ma tante Hélène et mon oncle Jean-Yves.

Résumé

Dans un contexte de vieillissement de la population des pays industrialisés se pose le problème de la prise en charge des personnes âgées dépendantes. La réponse proposée par les EPHAD ne pouvant constituer une réponse suffisante et acceptable pour les personnes faiblement dépendantes, la question de leur maintien à domicile dans un environnement sécurisé se pose.

Ainsi, le concept d'habitat intelligent pour la santé, issu de la domotique, a émergé dans les deux dernières décennies. Le principe général de ce domaine est de proposer un ensemble de solutions permettant la surveillance de l'observance, la détection précoce de maladies neurodégénératives, le maintien du lien social ou la détection de chutes et de situations d'urgences.

Dans cette idée et afin de s'affranchir de capteurs portés par l'habitant, nous avons proposé une méthode de détection des chutes basée sur la corrélation optique à partir de données issues de caméras vidéo. Notre approche se décompose en deux parties, l'une permettant l'identification de la personne présente sur l'image, l'autre le suivi de sa tête. L'étape de détection de la chute est réalisée à partir des données de vitesse verticale et horizontale.

Mots-clefs détection, suivi, corrélation, détection de chutes, reconnaissance, modèle linéaire

Abstract

In a context of a growing elderly population in industrialised countries, elderly-care is becoming a major problem. As traditional solutions (e.g. retirement homes) are no longer able to satisfy the increasing need, it would be desirable to give the healthiest part of this population a solution for their home care in a secure environment.

Thus, the concept of smart-home for health, coming from home automation, has emerged in the last two decades. The general principle of this field is to propose a set of solutions for monitoring compliance, early detection of neurodegenerative diseases, social interation assistance or fall and emergency situations detection.

Within this framework and to avoid the use of wearable sensors, we proposed a fall detection method based on optical correlation using video data. Our approach consists of two parts, one for identifying the person on the picture, the other for head tracking. The detection step is addressed by means of vertical and horizontal celerity.

Keywords detection, tracking, correlation, fall detection, recognition, linear model

Table des matières

Table des figures

Liste des tableaux

Acronymes

AUC : Area Under Curve
CPU : Central Processing Unit
EPHA : Établissements d'Hébergement pour les Personnes Âgées
FN : Faux Négatif
FP : Faux Positif
FPR : False Positive Rate
fps : frame per second
GPS : Global Positioning System
GPU : Graphics Processing Unit
JTC : Joint Transform Correlator
LM : Linear Model
MPEG : Moving Picture Expert Group
NL : Nonlinear
NZ : Non-Zero
PCE : Peak-to-Correlation Energy
PHPID : Pointing Head Pose Database
POF : Phase Only Filter
RAM : Random Access Memory
RFID : Radio Frequency Identification
ROC : Receiver Operating Characteristic
SNR : Signal-Noise Ratio
SoC : System on a Chip
TF : Transformée de Fourier
TN : True Negative
TP : True Positive
TPR : True Positive Rate
VLC : Vander Lugt Correlator
VN : Vrai Négatif
VP : Vrai Positif

Introduction

Introduction

Du fait de l'amélioration des conditions de vie, la pratique des mesures d'hygiènes et des progrès réalisés en médecine, la plupart des pays industrialisés connaissent un vieillissement de leur population – et, dans tous les cas, un rallongement de l'espérance de vie. Ainsi, dans le seul territoire français, la proportion des plus de 60 ans dans la population est passé de 19% (11 300 000 personnes) en 1992 à 23% (15 300 000) en 2012. Parallèlement, le nombre de places en EHPAD (Établissements d'Hébergement pour les Personnes Âgées Dépendantes) stagne, voire régresse : de 166 lits pour 1000 habitants de plus de 75 ans en 1996, on a atteint le nombre de 127 lits pour 1000 en 2006.

Ainsi, on assiste à un manque alarmant de moyens de prise en charge des personnes âgées, alors qu'elles se trouvent à un âge où elles sont particulièrement vulnérables et dépendantes. De plus, bon nombre de ces personnes ne souhaitent pas s'installer en maisons de retraite pour différentes raisons : pour conserver leur indépendance, leurs relations sociales, ou tout simplement pour rester dans un lieu familier – d'un point de vue médical, une perte de repères due à un changement d'habitat peut précipiter certaines maladies de vieillesse.

Le concept de l'habitat intelligent a ainsi émergé, adaptant les travaux de la domotique pour le bien-être, la détection précoce de pathologies ou la détection des chutes des personnes isolées, l'une des plus grandes causes de mortalité accidentelle de cette population. Les solutions actuellement commercialisées se présentent principalement sous la forme de capteurs portés par le patient, soit des accéléromètres, détectant automatiquement la chute, soit un bouton actionné par le patient lui-même. Or ces systèmes présentent divers problèmes. Le fait d'actionner un bouton d'alerte nécessite que le patient soit conscient, ou, dans tous les cas, dans la capacité de l'utiliser. La détection de chute par accéléromètre, quant à elle, est peu efficace en cas de chute lente, progressive. Enfin, ces deux solutions nécessitent que le patient porte le capteur au moment de la chute.

Afin de pallier ces problèmes, il est nécessaire d'imaginer un système capable d'interpréter une situation et de détecter et analyser un mouvement. Pour ce faire, nous proposons une surveillance automatique et autonome, entièrement intégrée à l'environnement. Un grand nombre de détecteurs installés dans la résidence de la personne dépendante permettrait de recueillir une large variété de types de données : audio, vidéo, infrarouge ou de pression (à partir de détecteurs embarqués dans le mobilier). Les données ainsi collectées seraient ainsi transmises à une unité de calculs locale pour l'expérimentation et l'analyse.

Ainsi, il serait possible de considérer une kyrielle de situations telles que des chutes, une inaction anormale ou un changement brutal dans les habitudes. Les informations obtenues sur ces événements seraient finalement transmises aux services de secours et pourraient de plus fournir une aide au diagnostic. Finalement, une alerte pourrait être envoyée par SMS (Short Message Service) ou courrier électronique aux proches ou aux voisins de la personne.

L'objectif des travaux présentés dans ce manuscrit et réalisés en collaboration avec la mutuelle de santé Malakoff-Médéric et la société informatique Open est de répondre à ces exigences et de proposer un prototype d'un système de détection des chutes des personnes âgées. Ce système devra être intégré à l'habitat (s'affranchissant de port de capteurs) et devra être capable d'intégrer différentes fonctionnalités supplémentaires telles que la détection de comportements pathologiques (à l'aide d'un apprentissage des habitudes du patient).

Ce manuscrit de thèse est organisé en trois parties principales.

La première partie est dédiée au contexte dans lequel s'inscrivent les travaux présentés dans cette thèse. Dans le premier chapitre, nous nous intéressons au contexte général. En effet, nous présentons tout d'abord les problèmes de prise en charge soulevés par le vieillissement de la population. Cette analyse nous entraîne vers l'émergence des dispositifs de maintient à domicile des personnes âgées et plus particulièrement des systèmes de détection de chute, l'objet de cette thèse. Les capteurs utilisés ainsi que les système de détection de chute par vidéo-surveillance sont développés. Finalement, notre système est présenté ainsi que les besoins énoncés par nos partenaires Malakoff-Médérik et Open. Dans le second chapitre, nous nous focalisons sur les méthodes de corrélation, organisées en deux familles d'architecture : le corrélateur de Vander-Lugt et le corrélateur à spectre joint. Après avoir présenté les critères d'évaluation de la corrélation, nous soulignons la pertinence de la corrélation numérique. Enfin, les deux architectures sont présentées et discutées dans un soucis d'application à un système permettant le suivi et l'identification.

La seconde partie de ce mémoire, composée de deux chapitres, est consacrée à l'utilisation de la corrélation pour l'identification et le suivi. Le premier chapitre expose nos travaux sur le corrélateur de Vander-Lugt pour l'identification. Tout d'abord nous présentons notre méthode de débruitage du plan de corrélation à l'aide d'un modèle linéaire. Les différents paramètres du modèle sont ensuite étudiés. Enfin nous terminons ce chapitre par une évaluation de notre approche sur une base de données de référence. Dans le second chapitre, nous nous focalisons à l'utilisation du corrélateur à spectre joint pour le suivi. Nous commençons par une brève description des capacités de localisation du corrélateur à spectre joint pour continuer sur l'application de cette caractéristique au suivi d'objets dans une séquence vidéo. Ainsi, nous présentons le principe général de notre algorithme de suivi. Un protocole d'expérimentation de ses performances en fonction des nombreux paramètres disponibles est proposé et appliqué à notre méthode. Enfin, nous détaillons les optimisations apportées à notre approche afin notamment de lui permettre une ré-initialisation automatique en cas de mauvaise détection.

La dernière partie, quant à elle, expose l'application des travaux présentés dans la partie précédente à un système de détection des chutes de la personne âgée dépendante. Nous commençons par détailler notre système de détection de chute ainsi que notre installation pour l'expérimentation. Dans un second temps nous explorons notre méthode d'identification basée sur le débruitage du plan de corrélation sur une base de données prises en conditions réelles. Les résultats obtenus sont comparés avec des méthodes de la littérature. Subséquemment nous appliquons notre algorithme de suivi à la détection des chutes. Pour cela, nous commençons par exposer notre base de données, réalisée dans un soucis d'analyse des conditions limites de notre algorithme et de mise en condition dans un scénario de suivi et de détection des chutes de la personne. Les résultats sur l'ensemble de ces séquences sont ensuite présentés. Enfin, nous appliquons à notre système un critère de détection des chutes basé sur la vitesse horizontale et verticale de la personne.

Ce manuscrit s'achève sur une synthèse des travaux effectués ainsi que sur les perspectives offertes par ce travail de thèse.

Première partie

Contexte de la thèse

Chapitre 1

Contexte et état de l'art

Sommaire

Dans un contexte de vieillissement de la population, nous sommes aujourd'hui devant un besoin grandissant de prise en charge de la personne âgée faiblement dépendante. Les solutions existantes se présentent comme un ensemble de résidences ou d'établissements médicalisés. Malheureusement, ces infrastructures représentent généralement un coût élevé pour la personne et sa famille. De plus, le manque criant de places engendre une attente prolongée. Enfin, la perte de repères engendrée par la modification radicale de la vie quotidienne peut accélérer l'apparition et l'évolution de maladies neuro-dégénératives.

Pour pallier à l'ensemble de ces inconvénients, différentes solutions d'habitat intelligent commencent à apparaître et ce domaine, à l'origine un cas particulier de la domotique, est devenu un secteur en pleine expansion de la recherche scientifique [1].

Dans ce chapitre, nous présentons tout d'abord le contexte historique dans lequel s'inscrit ce travail de thèse. Ainsi, un aperçu de l'ampleur du vieillissement de la population générale et du besoin impérieux de solutions adaptées sera exposé. En second lieu, nous présentons les systèmes existants de détection de chute suivant les différents capteurs utilisés. Puis nous décrivons et discutons des méthodes présentes dans la littérature pour la détection des chutes par vidéo-surveillance. Enfin, nous introduisons notre méthode de détection des chutes, en partenariat avec Malakoff-Médérik et Open.

1.1 Contexte

1.1.1 Une population vieillissante

La plupart des pays industrialisés connaissent actuellement un vieillissement prononcé de leur population. Cela est dû à différents facteurs, l'augmentation de l'espérance de vie, une baisse de la natalité et les conséquences du pic de naissances de l'après guerre, appelé "Baby boom". Les prévisions pour l'Europe de l'Ouest, par exemple, sont d'une augmentation de 20% en 2000 à 42% en 2050 de la proportion de personnes âgées de plus de 60 ans par rapport à la population totale [2]. En ce qui concerne la France, la projection est un vieillissement de la population jusque 2050 voire 2060 [3]. L'évolution de la pyramide des âges est illustrée en figure 1.1. Nous observons tout d'abord sur les courbes tirées des chiffres de 2007 trois creux dans la population, aux alentours des personnes âgées de 90 ans, de celles âgées de 65 ans et enfin de celles âgées 30 ans. Le premier correspond au déficit des naissances dû à la première guerre mondiale, le second au déficit des naîssances dû à la seconde guerre mondiale. Suite à cela une augmentation importante de la population est visualisable sur la courbe, se terminant au troisième creux. Il s'agit de l'augmentation des naîssances faisant suite à la guerre, le "Baby boom". En 2007, cette population avait moins de 60 ans. Or en 2060 celle-ci formera la population âgée c'est-à-dire de plus de 65 ans. Comme nous pouvons le constater dans la pyramide des âges, cette entrée dans l'âge avancé de la génération du "Baby boom" entraine un décuplement de la population dite "âgée". On estime ainsi que 29% de la population aura plus de 65 ans en 2050 contre 16% en 2000. Le nombre de plus de 75 ans sera multiplié par trois, de plus de 85 ans par quatre.

1.1.2 La prise en charge

L'âge entraîne un certain nombre de réductions des capacités physiques, entraînant un besoin d'assistance pour certains ou la majorité des actes de la vie quotidienne selon le degré de dépendance. Une personne âgée dépendante est définie par : "Toute personne d'au moins soixante ans (âge à partir duquel selon la loi on peut prétendre à l'Allocation Personnalisée d'Autonomie) qui, nonobstant les soins qu'elle est susceptible de recevoir, a besoin d'une aide pour l'accomplissement des actes essentiels de la vie ou dont l'étant nécessite une surveillance régulière" [4].

Ainsi on estime à 136 000 les personnes dépendantes vivant seules à domicile en France, dont la moitié doit être aidée pour la réalisation de tâches comme le lever du lit ou l'habillage. Cette population est particulièrement à risque pour les accidents domestiques, étant dans l'incapacité de se relever après une chute.

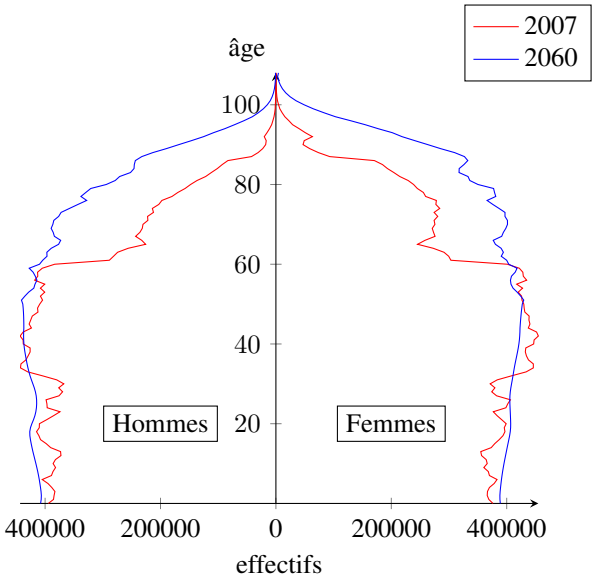


FIGURE 1.1 – Pyramide des âges au premier janvier 2007 et prédictions pour 2060 [3].

Nous reproduisons en tableau 1.1 le nombre d'années moyen de vie en situation de dépendance avant d'être admis dans un établissement spécialisé. On observe ainsi que la durée moyenne de vie dans la famille d'une personne ayant des difficultés pour se lever est de 3 ans pour les femmes et 5 ans pour les hommes. Si une telle personne vit en indépendance, elle attend en moyenne 1,5 ans. Également, une femme devant rester confinée dans son lit restera 3,5 ans dans sa famille. De telles situations augmentent considérablement les risques d'accidents domestiques graves et de perte de lien social. En effet, l'état physique de ces personnes rendent le risque de chute important. De plus, l'intervention d'un tiers est indispensable pour lui permettre de se relever. Évidemment, ces chiffres sont à relativiser, ne permettant pas de dissocier les personnes ayant prononcé le souhait de vivre dans une telle situation. Cependant, elles traduisent un manque abyssal de places disponibles en établissement et elles dévoilent en outre le besoin de proposer une solution pour permettre à ces personnes de vivre tout en limitant les risques inhérents à leur âge avancé.

1.2 Les systèmes de détection de chute

Le vieillissement de la population des pays industrialisés, de part les nombreux problèmes de santés engendrés par un âge avancé, entraîne un besoin de nouvelles solutions pouvant permettre le maintien à domicile de la personne faiblement dépendante en sécurité.

Ainsi, et de part l'arrivée de capteurs et d'unités de calcul de faible coût, le concept "d'habitat intelligent pour la santé" a rapidement émergé et est devenu un domaine florissant [1, 6, 7, 8]. Demiris et al. [1] décrit le terme "habitat intelligent" comme "une résidence équipée de technologies facilitant l'observation de résidents et/ou améliorant l'indépendance et la qualité de vie". Plus précisément, le terme d'"habitat intelligent pour la santé" s'inscrit dans le cadre de technologies permettant :

— l'observation des données physiologiques,
— l'observation des données fonctionnelles,
— la détection et la réponse aux situations d'urgence,
— la prévention des accidents domestiques,

TABLE 1.1 – Durées moyennes vécues avec les incapacités avant d'entrer dans l'établissement habité à la date de l'enquête, en années [5].

Incapaticés	Domicile précédent des hommes			Domicile précédent des femmes		
	Ordinaire indépendant	Famille	Autre institution	Ordinaire indépendant	Famille	Autre institution
Toilette, habillage, alimentation						
Difficultés pour la toilette	2,0	8,5	7,5	2,0	4,0	2,5
Difficultés pour l'habillage	2,5	8,0	6,0	1,5	4,0	2,5
Difficultés pour couper la nourriture et se servir à boire	2,5	8,0	7,0	2,0	2,5	3,0
Difficultés pour manger (une fois le repas prêt)	1,5	11,5	6,0	1,0	3,5	2,0
Elimination						
Difficultés pour aller aux toilettes	1,5	5,5	5,0	1,5	3,0	2,5
Difficultés pour contrôler selles et urines	2,0	1,0	4,5	1,5	1,5	2,5
Mobilité						
Doit rester confiné à l'intérieur (lit, chambre, institution)	2,0	6,5	4,0	1,5	3,5	2,0
Difficultés pour se coucher, se lever, s'assoir	1,5	5,0	4,5	1,5	3,0	1,5
Difficultés pour se déplacer au même étage	2,0	3,5	3,0	1,5	1,0	1,5
Communication, cohérence, orientation						
Difficultés pour communiquer	3,0	8,0	10,5	2,0	5,5	2,5
Difficultés pour se souvenir du moment de la journée	2,0	9,0	6,5	1,5	3,0	2,5
Sens						
Difficultés pour voir de près	4,5	6,0	6,0	5,0	6,5	5,0
Difficultés pour voir de loin	4,5	1,0	2,5	3,0	5,5	2,5
Difficultés pour suivre une conversation	6,5	6,5	7,0	3,5	3,5	3,5
Difficultés pour parler	5,5	12,0	13,0	2,5	8,0	3,5
Souplesse						
Difficultés pour se servir de ses mains et doigts	3,5	9,5	7,0	2,0	3,5	2,5

— la préservation du lien social,
— l'assistance cognitive.

Il s'agit donc d'un large ensemble d'applications permettant d'assister la personne dépendante dans sa vie quotidienne et de prévenir ou de répondre aux risques induits par la vie en situation isolée. Ces systèmes ont par conséquent un champ d'action aussi large que la prévention de la démence, la délivrance ou le rappel automatique de la thérapie, la détection de chutes ou encore la proposition d'appareils électroménagers adaptés, par exemple.

La chute accidentelle constitue l'un des dangers les plus fréquents et les plus sérieux pour une personne dépendante isolée. En effet, on estime à environ un tiers la proportion des plus de 65 ans effectuant au moins une chute annuelle [9]. Celles-ci peuvent engendrer des conséquences sévères, soit directement (fractures ou autres traumatismes), soit indirectement (baisse des capacités fonctionnelles). Elles représentent la cinquième cause principale de décès dans la population âgée [10]. Dans la littérature, différents travaux d'état de l'art ont été réalisés sur les systèmes d'"habitat intelligent" ou plus précisément sur des applications de détection des chutes de l'habitant. Ces dispositifs peuvent permettre soit la prévention de la chute, en assistant la personne dans sa vie quotidienne et en apprenant son comportement, soit la détection de la chute et le déclenchement d'un signal d'alerte. Nos travaux se situent dans cette seconde voie. Le lecteur pourra consulter les publications suivantes : [1, 6, 7, 8, 11, 12, 13, 14]. Une chute est une situation complexe et variée, que nous caractérisons dans le paragraphe suivant.

1.2.1 La définition de la chute

Afin d'être en mesure de réaliser un système de détection de la chute, il est impératif de savoir la définir. Ainsi, Noury et al. [15] ont caractérisé la chute (Fig. 1.2) en une une décomposition en 4 sous-événements :

Phase de "pré-chute" : Il s'agit de la période précédant la chute, où la personne effectue ses activités de la vie quotidienne. Elle comprend certains mouvements soudains tels que "s'asseoir" ou "s'accroupir"

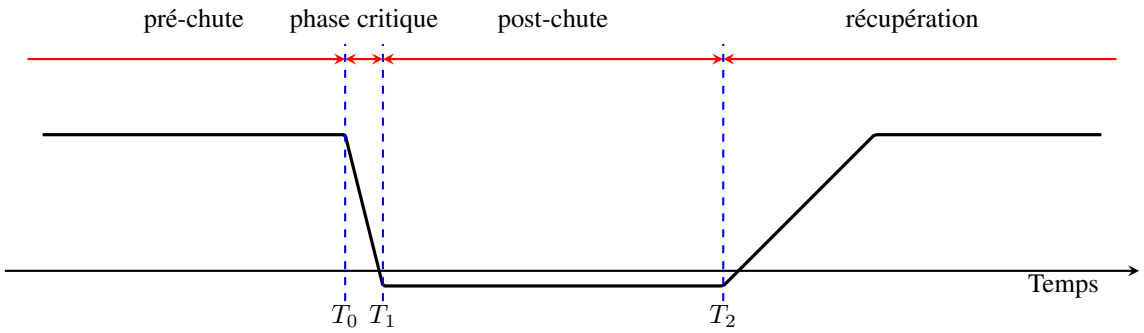


FIGURE 1.2 – Les quatres phases d'une chute [15].

ne devant pas engendrer de fausses alarmes [1].

Phase critique : Cette période correspond à la chute à proprement parler. Elle dure un temps très court (entre $300ms$ et $500ms$) et se termine par un choc sur le sol.

Phase de "post-chute" : La personne est allongée sur le sol.

Phase de récupération : La personne peut éventuellement se relever seule ou à l'aide d'un tiers.

D'après Rougier [16], les méthodes de détection de chute peuvent donc être classées suivant la phase de la chute sur laquelle elles se focalisent. Ainsi, les méthodes directes sont basées sur une détection de la phase critique (e.g. vitesse du corps, impact au sol) et les méthodes indirectes sur une détection de la phase "post-chute" (e.g. posture, inactivité).

L'événement "chute" peut varier sensiblement selon la situation. En effet, une chute à partir d'une position assise, par exemple, ne se fera ni à la même vitesse, ni à la même hauteur qu'une chute ayant lieu alors que la personne est en déplacement. Pour répondre à ce problème, Yu [13] a proposé une description des caractéristiques de chute suivant trois situations différentes lors de la phase de "pré-chute" :

Les caractéristiques d'une chute à partir d'un lit :

- Une chute dure de 1 à 3 secondes et est composée de plusieurs sous-actions,
- la personne est allongée sur le lit au début de la chute,
- le corps réduit sa hauteur depuis celle du lit jusqu'à celle du sol (pendant cette période, le corps chute librement),
- le corps est près du lit à la fin de la chute.

Les caractéristiques d'une chute à partir d'une chaise :

- Une chute dure de 1 à 3 secondes et est composée de plusieurs sous-actions,
- la personne est assise sur la chaise au début de la chute,
- la tête réduit sa hauteur depuis celle à la position assise jusqu'à celle du sol (pendant cette période, la tête chute librement),
- le corps est près de la chaise à la fin de la chute.

Les caractéristiques d'une chute à partir de l'orthostatisme :

- Une chute dure de 1 à 2 secondes et est composée de plusieurs sous-actions,
- la personne est debout au début de la chute,
- la tête est sur le sol à la fin de la chute (celle-ci peut avoir un faible mouvement durant cette phase),
- une personne chute brusquement dans une direction,

1. Détection d'une chute inexistante

— la tête réduit sa hauteur depuis celle à la station debout jusqu'à celle du sol (pendant cette période, la tête chute librement),
— la tête est localisée à la fin de la chute autour d'un cercle dont le centre est la position des pieds avant la chute et dont le rayon est la taille de la personne.

Enfin, une étude de la chute et de sa distinction des activités de la vie quotidienne a été réalisée par Wu en 2000 [17]. Son travail se focalise sur la vitesse horizontale et verticale lors d'activités effectuées par 3 sujets. Ils mettent en avant le fait que la vitesse horizontale et verticale sont multipliées simultanément par trois dans un laps de temps compris entre $300ms$ et $400ms$ avant la fin de la chute. À l'inverse, les activités courantes de la vie quotidienne étudiées dans cet article (i.e. marches, s'asseoir, se lever, descendre des marches, attraper un objet, s'allonger sur un lit, entrer et sortir d'une baignoire) présentent toutes des vitesses verticales et horizontales limitées (entre 0 et $1500mm.s^{-1}$) et évoluant indépendemment.

Ainsi, une chute est un événement décomposé en 4 sous-parties (phases "pré-chute", critique, "post-chute" et de récupération). La grande variété de situations lors de la phase "pré-chute" entraine une variabilité de certains paramètres tels que la hauteur de chute, la durée de la phase critique. Cependant, l'augmentation importante des vitesses horizontale et verticale de la tête, ainsi que la position du corps lors de la phase "post-chute" sont des constantes de cet événement. La prise en compte de ces caractéristiques est indispensable quant à la détection de chutes et la réduction du nombre de fausses alarmes, un des problèmes majeurs de ce type d'installations. En outre, l'observation de la phase critique est la voie offrant la plus grande capacité d'analyse, phase sur laquelle nos travaux se focalisent.

1.2.2 Les capteurs utilisés dans la détection des chutes

Comme nous venons de le voir, la chute est une action complexe, avec une grande variabilité des caractéristiques suivant les situations. Celles-ci différeront sensiblement selon la phase de "pré-chute" ou selon qu'il s'agit d'une chute accidentelle ou syncopale. En effet, une chute accidentelle peut être beaucoup plus brutale qu'une chute syncopale. À l'inverse, une chute syncopale est caractérisée par l'absence de la capacité de la personne de déclencher l'alerte et de se protéger lors de l'impact. Différents capteurs peuvent être utilisés pour la détection de la chute, soit lors de la phase critique, soit lors de la phase de "post-chute". Ceux-ci peuvent être dédiés à la détection de la chute, la réduction du nombre de fausses alarmes ou la collecte d'informations sur la personne pour améliorer la prise en charge ou les performances du système (e.g. capteurs de présence).

Noury et al. [12], Yu [13] et Mubashir et al. [14] ont compilé un grand nombre de systèmes de détection de chutes et effectué une classification de ces méthodes. S'inspirant de [12] et [14], nous effectuons une classification des capteurs. Nous choisissons une séparation de ces systèmes en deux classes distinctes (Fig. 1.3) suivant le type de capteur utilisé : capteur porté, ou capteur environnemental. Un troisième type de capteur peut être utilisé en "habitat intelligent", les capteurs de médiation de l'infrastructure [18] (capteurs installés sur les appareils de la vie quotidienne : électroménager, compteur d'eau, chauffage, prises, etc), mais leur pertinence quant à la détection de la chute est minime.

1.2.2.1 Les capteurs portés

De part leur fiabilité, leur simplicité d'implantation et leur faible coût, les capteurs portés sont utilisés dans un grand nombre de systèmes de détection de chutes. Ils peuvent se présenter sous la forme de boutons poussoir [19, 20], d'accéléromètres [20, 21, 22, 23, 24, 25], de gyroscopes [25, 26], de marqueurs RFID (Radio Frequency Identification) [21, 22] ou de puces GPS (Global Positioning System) [27]. Ces capteurs peuvent aussi être utiles pour collecter des données complémentaires.

Les systèmes actuellement commercialisés utilisent principalement un bouton poussoir porté en permanence par l'habitant et actionné manuellement en cas d'urgence [19]. Ce type de solution présente l'avantage d'être facilement et rapidement déployable à grande échelle et de générer peu de fausses alarmes de part

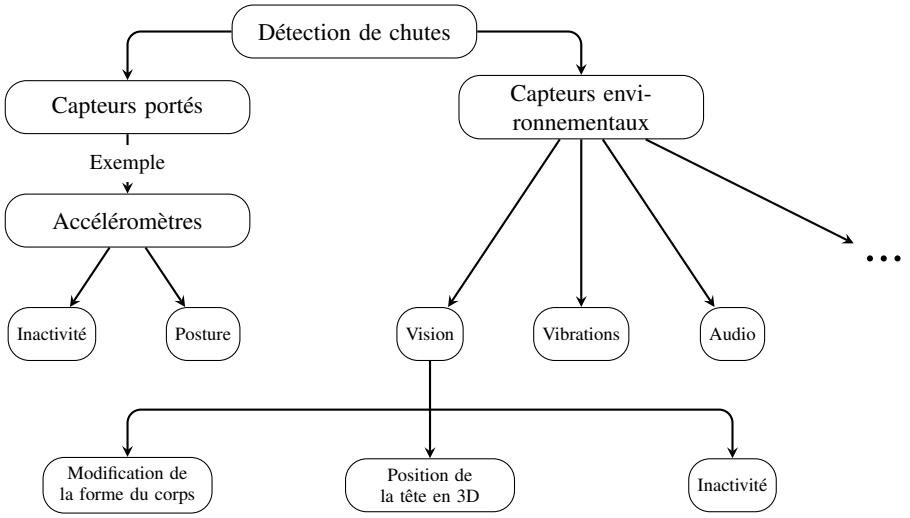


FIGURE 1.3 – Classification des méthodes de détection de chutes.

l'absence de besoin d'une analyse des données collectées. Néanmoins, la personne peut se retrouver dans l'incapacité d'actionner le système en cas de perte de conscience ou de blessure. Pour remédier à cela, d'autres solutions ont été commercialisées, comme le "Tunstall Telecare" [28], basées sur l'analyse automatique de l'information extraite par un accéléromètre porté par le patient.

Särelä et al. [20] proposent un système contenant un bracelet et une station centrale, appelée "IST VIVAGO". Au moyen d'un accéléromètre, ils sont à même de détecter une chute et d'analyser le rythme circadien de la personne, complétant l'information disponible au personnel de soin. L'utilisation d'accéléromètres fixés ou directement intégrés dans le téléphone portable a également été explorée en 2006 par Zhang et al. [29] et en 2010 par Dai et al. [30]. Néanmoins, l'une des principales limitations des accéléromètres est leur incapacité à détecter des chutes lentes, suite à une perte de connaissance par exemple.

Également, certains systèmes utilisent la technologie RFID [21, 22] en complément d'une méthode de détection des chutes pour la localisation et l'identification de la personne. Elle présente le double avantage de sa simplicité d'utilisation et son très faible coût. Un système RFID est composé de trois sous-parties : (i) le marqueur ; (ii) le lecteur ; (iii) le système de traitement. En complément d'un message d'identification, le signal RSS (Received Signal Strength), permettant la mesure de la puissance du signal, permet la localisation du marqueur [31]. Deux types de marqueurs sont disponibles, les marqueurs passifs (rayon d'action de $\sim 1m$) et les marqueurs actifs (rayon d'action $\sim 300m$). Outre leur large rayon d'action, les marqueurs actifs ont la capacité d'envoyer des messages d'identification. Ainsi, la "Smart Home Care Network" [21, 22] utilise un marqueur RFID pour localiser la personne à terre, après la détection d'une chute à l'aide d'un accéléromètre. Lorsqu'une chute est détectée, une caméra vidéo effectue une analyse de posture, afin de réduire le taux de fausses alarmes. Leur système permet de détecter le patient dans 90% des images. Malheureusement, cette approche est validée sur un ensemble de seulement 16 images qui est insuffisante pour couvrir l'ensemble des situations possibles et les problèmes environnementaux.

Dans le système proposé par Demongeot et al. [23], appelée "Health Integrated Smart Home Information System", accéléromètres, capteurs infrarouges et capteurs de porte magnétiques sont utilisés. Comme avec les systèmes décrits précédemment, cette approche présente l'inconvénient majeur d'être basée sur des capteurs portés.

Des gyroscopes ont également été utilisés par Bourke et al. [26] ou par Nyan et al. [25] en complément d'accéléromètres afin de déterminer l'orientation du corps.

Une fusion de capteurs portés et environnementaux a été apportée par Cavalcante Aguilar et al. en 2011 [32]. Leur méthode permet un taux de bonne détection de 95% pour 5% de fausses alarmes [33].

Finalement, Fleck et al. [27] proposent l'utilisation combinée de caméras pour la détection de la chute et d'une puce GPS pour la localisation.

Les capteurs portés permettent une simplicité d'utilisation et de traitement des données. En outre, il est aisé de les compléter par un dispositif de mesure des données physiologiques. Néanmoins, ils souffrent de certains défauts majeurs : ils peuvent être oubliés par l'habitant, le blesser lors de l'impact et enfin nécessitent un remplacement régulier de la batterie.

1.2.2.2 Les capteurs environnementaux

Une grande variété de capteurs peuvent être installés directement dans l'habitat de la personne dépendante : microphones, caméras vidéo, capteurs de pression, capteurs de vibrations, actionneurs de porte, caméras infrarouge.

Ainsi, en 2005, Helal et al. [34] proposent un espace d'expérimentation sous la forme d'une maison entièrement équipée pour la prise en charge de la personne dépendante, comprenant notamment un sol composé de capteurs de pression, capable de localiser les occupants et dont les données peuvent permettre la détection d'éventuelles chutes. Également, [35] proposent l'utilisation de capteurs de sol pour la détection des chutes. Le problème majeur de cette solution est qu'elle est intéressante pour la construction de futurs logements dédiés à une personne dépendante mais qu'elle est peu appropriée pour une installation dans des logements existants. Également, Nishida et al. [36] utilisent un lit doté de capteurs de pression pour détecter la présence de la personne et donc son éventuelle chute ainsi que des microphones.

Également, des capteurs vibrationnels peuvent être utilisés pour analyser l'impact de la chute sur le sol [37, 38]. Cette approche permet de détecter efficacement les chutes brutales mais ne permet pas la détection des chutes molles. De plus, la grande variabilité des sols (moquette, parquet, carrelage) impacte fortement les performances de la méthode.

Finalement, un grand nombre de systèmes, décrits précisément en partie 1.3 page 15, utilisent les informations issues de caméras vidéo ou infrarouge pour la détection de la chute [39, 21, 22, 40, 34, 41], nécessitant des méthodes de traitement d'images. Quelques systèmes basés sur des caméras vidéo ont été commercialisés. Le projet EDAO [42] utilise différents capteurs (caméras vidéo, des capteurs infrarouge, des détecteurs de porte) pour la détection de posture et de zones interdites dans la pièce. Lorsqu'une alerte est générée, l'image de la scène est alors envoyée à un opérateur qui décide de l'urgence de la situation. Ce système présente l'inconvénient de générer un grand nombre de fausses alarmes nécessitant l'intervention d'un opérateur et le rendant par conséquent très invasif.

De Silva et al. [41] combine des informations de micros et de caméras vidéo pour détecter les situations d'urgence (chutes, omission de thérapie, cris). À cet effet, la détection est effectuée par le patient au moyen d'une soustraction de fond. L'arrière-plan est mis à jour à chaque trame en fonction de l'emplacement de la personne. Un modèle de forme / épaule est ensuite appliqué sur l'objet, le tronc correspondant étant identifié à travers une procédure d'appariement d'histogrammes. Le processus de suivi est réalisé avec un appariement de l'histogramme. Enfin, la reconnaissance d'événements audio est effectuée par une détection du contour du signal. La précision de détection de chute à l'aide de la méthode vidéo et audio est de $93{,}3\%$ et $91{,}67\%$, respectivement. Aucune information n'est disponible sur la manière dont ils fusionnent les données extraites de capteurs audio et vidéo.

De même, Rougier et al. [43] proposent en 2011 l'utilisation de caméras Kinect [44] pour la réalisation de cartes de distance. En effet, le capteur Kinect est doté d'un projecteur de mire infrarouge permettant l'estimation robuste et rapide de la distance. Néanmoins, ce capteur présente l'inconvénient d'être soumis à une licence ne permettant pas la commercialisation.

Les capteurs environnementaux ont le double avantage d'être inclus dans l'environnement et de permettre la collecte de données riches en informations. Ce dernier avantage représente également leur faiblesse. En

effet, ils nécessitent un traitement plus lourd des données, la réalisation d'un système performant est donc plus complexe.

1.2.3 Discussion et conclusion

Un large éventail de capteurs est disponible pour les applications de détection des chutes de la personne. Nous reproduisons sur le tableau 1.2 un résumé des principales caractéristiques de quelques-uns des capteurs utilisés dans la littérature.

Les systèmes portés sont des solutions efficaces par leur faible coût et leur simplicité d'installation et de développement. En effet, l'utilisation d'un bouton poussoir permet la réalisation et la commercialisation rapide et à un coût relativement faible d'un système de gestion des situations d'urgence. Ils ont l'indéniable avantage de ne générer que peu de fausses alarmes et de pouvoir être déclenché à tout moment par l'habitant. Cependant, leur activation nécessite une action effectuée par la personne pendant une situation d'urgence. Cela est particulièrement problématique, le sujet pouvant être inconscient ou dans l'incapacité de réaliser un mouvement de par les conséquences de l'impact sur le sol. Ce type de capteur est donc une solution intéressante à court terme ou en complément d'autres capteurs. Les accéléromètres et les gyroscopes quant à eux sont utilisés dans un très grand nombre de systèmes. Contrairement aux systèmes utilisant un bouton poussoir, ils ont la capacité de collecter des données permettant la détection automatique de la chute. Leur tendance à être implantés directement dans les téléphones portables ou dans des montres bracelets les rend particulièrement intéressants pour un déploiement extrêmement rapide. Malheureusement, le nombre élevé de fausses alarmes engendrés par les accéléromètres rend nécessaire leur utilisation simultanément avec d'autres capteurs complémentaires (e.g. vision).

En ce qui concerne les capteurs environnementaux, ils sont intégrés dans l'ensemble de l'habitation sous la forme de capteurs binaires ou caméras par exemple. Leur installation et maintenance peuvent s'avérer coûteuses et des problèmes supplémentaires apparaissent lors de la présence simultanée de plusieurs résidents. Les capteurs binaires se présentent généralement sous la forme de détecteurs de mouvement ou de pression et ont une utilisation relativement limitée, principalement de complément d'informations provenant d'autres capteurs (e.g.caméras vidéo). Ils sont peu coûteux, simples d'installation et de traitement et ne génèrent pas de problème éthique. En revanche, les caméras vidéo sont des sources d'information très riches mais de traitement plus complexe. En outre, elles présentent l'avantage de pouvoir être étendues à un système plus complet, prenant en compte notamment les habitudes de la personne pour la prévention et la détection des chutes. En ce qui concerne les systèmes vibrationnels, ils représentent une solution efficace pour la détection de chutes brutales mais doivent être complétés par d'autres capteurs (e.g. accéléromètres) et leur traitement est dépendant des caractéristiques du sol de l'habitation. Finalement, les capteurs de pression positionnés dans le dallage (Smart Floor) nécessitent une installation lourde et coûteuse réduisant de fait leur pertinence dans un objectif de maintien à domicile de la personne.

Pour s'affranchir de ces limitations et afin de prendre en compte les avantages proposés par l'information riche obtenue par la vision et sa faible intrusivité, un grand nombre de systèmes basés sur des caméras vidéos ont été proposés.

1.3 Systèmes de détection de chute basés sur la vision

La détection de chute par vidéo-surveillance a bénéficié d'un grand intérêt cette dernière décennie, notamment grâce à la mise sur le marché à un coût abordable d'unités de calcul performantes. Ces systèmes peuvent être classés suivant différentes catégories suivant la méthode employée pour l'analyse de la scène et la détection de la chute : (i) l'analyse de l'inactivité, se focalisant sur la phase de "post-chute" ; (ii) la détection de posture, pouvant être appliquée suivant la méthode sur la phase critique ou la phase de "post-chute" ; (iii) le suivi de la tête de la personne, permettant l'analyse de l'ensemble des quatre phases de la chute.

TABLE 1.2 – Comparaison non-exhaustive des capteurs utilisés pour la détection des chutes.

Approche	Capteur	Coût	Intrusif	Précision	Installation	Robuste	Publication
Capteurs portés	Bouton poussoir	Faible	✓	Dépendante du scénario	Simple	×	Life Alert [19] Särelä et al. [20]
	Accéléromètre	Faible	✓	Dépendante du scénario	Simple	×	Särelä et al. [20] Tabar et al. [21] Keshavarz et al. [22] Demongeot et al. [23] Estudillo-Valderrama et al. [24] Nyan et al. [25] Tunstall Telecare [28] Zhang et al. [29] Dai et al. [30]
	Gyroscopes	Faible	✓	Elevée	Simple	✓	Demongeot et al. [23] Bourke et al. [26] Nyan et al. [25]
Capteurs environnementaux	Microphones	Faible	×	Dépendant du scénario	Moyenne	×	Chahuara 2011 [45] Nishida et al. [36] De Silva et al. [41] Tabar et al. [21]
	Vibrations	Faible	×	Faible	Moyenne	×	Alwan et al. [37] Zigel et al. [38]
	Smart floor	Elevé	×	Moyenne	Difficile	Moyenne	Helal et al. [34] Nishida et al. [36]
	Caméra vidéo	Faible à moyen	×	Elevée	Moyenne	✓	Williams et al. [39] Tabar et al. [21] Keshavarz et al. [22] Aghajan et al. [40] Helal et al. [34] De Silva et al. [41] Auvinet et al. [46]

1.3.1 Systèmes basés sur l'inactivité

Comme nous l'avons explicité en partie 1.2.1 page 10, une chute se termine généralement par une phase d'immobilité de la personne sur le sol. Les systèmes basés sur l'analyse de l'inactivité exploitent cette caractéristique de la chute.

Ainsi, McKenna et Nait-Charif [47] proposent une méthode de détection des chutes utilisant une unique caméra grand-angle fixée au plafond de la pièce. Aucune rectification ou calibration n'est nécessaire. Leur système est basé sur une modélisation du contexte spatial, caractérisé en différentes zones : (i) des zones d'entrée dans la pièce ; (ii) des zones d'inactivité ; (iii) des zones de transition. Les zones d'inactivité correspondent à des régions où la personne est susceptible de rester immobile (e.g. fauteuil). L'apprentissage de ces zones est effectué à l'aide d'une estimation de mixtures de gaussiennes bayesiennes. Quant au suivi de la personne, il est réalisé à l'aide d'un filtrage particulaire à partir d'une détection par soustraction de fond. La personne est représentée par une ellipse.

Jansen et Deklerck [48], quant à eux, s'inspirent de McKenna et Nait-Charif et se basent sur l'information 3D obtenue à l'aide d'une caméra TOF (Time of Flight) pour la détection d'inactivités. Des images 3D, ils extraient la localisation 3D et l'orientation de la personne, leur permettant de déterminer deux seuils de détection de chute : (i) la localisation est utilisée pour la détection d'inactivité ; (ii) l'orientation est utilisée pour détecter une personne à terre. Finalement, un modèle contextuel est mis en place, prenant en compte l'inactivité et l'orientation afin de déterminer s'il s'agit d'une situation d'urgence.

L'analyse de l'inactivité a l'avantage de ne nécessiter qu'une faible analyse du contexte, notamment des phases "pré-chute" et critique. Néanmoins, il s'agit également de sa principale limitation, ne permettant pas d'analyser les circonstances de la chute. En outre, l'utilisation de l'analyse d'autres phases de la chute est nécessaire pour la réduction du nombre de fausses alarmes.

1.3.2 Systèmes basés sur la détection de posture et le changement de la forme

Un grand nombre de systèmes de détection de chute utilisent la détection de posture de la personne. En effet, cette approche est relativement bien adaptée pour l'implantation de systèmes bas coût, de par sa capacité à détecter une chute en utilisant une fréquence de rafraîchissement de l'image très faible.

Tao et al. [49] développent en 2005 un système permettant la détection et le suivi d'une personne, complété par une méthode de détection de chute basé sur un rapport hauteur/largeur. La détection de la personne est effectuée à l'aide d'une soustraction de fond et une opération d'ouverture pour la réduction du bruit. La soustraction de fond est réalisée dans l'espace de couleurs HSV (Hue-Saturation-Value), chaque pixel est représenté par une mixture de gaussiennes. Pour ce qui est de la détection des chutes, le rapport entre la hauteur et la largeur de la personne autorise une classification des situations entre une personne à terre, assise, ou en activité de marche. Une comparaison a été effectuée entre l'utilisation d'une caméra positionnée à la hauteur d'une table et une caméra au plafond. Les performances de l'étape de détection de chute sont largement détériorées lors de l'utilisation d'une caméra au plafond, bien que cette disposition permette d'éviter les sorties de champ. D'après les auteurs, ceci est certainement induit par l'angle de dépression de la caméra, réduisant les modifications du rapport hauteur/largeur.

De la même façon, Miaou et al. [50] utilisent l'information issue d'un "omni-caméra". Pour ce faire, un miroir convexe fixé au plafond est filmé en permanence par une caméra CCD standard. Ce type d'installation, appelée "MapCam" permet d'obtenir simplement des images à 360° et ainsi éliminer les cas de hors champ. Leur système est basé sur une détection de la silhouette de la personne à l'aide d'une soustraction de fond. Une opération morphologique d'ouverture est appliquée afin de réduire le bruit présent à l'issue de l'étape soustraction de fond. Tout comme le système de Tao et al. [49], la chute est détectée à partir d'un simple rapport entre la largeur et la hauteur de la personne détectée. Si ce rapport est supérieur à un seuil fixé, un signal d'alarme est envoyé. Enfin, une utilisation d'informations personnelles comme la hauteur, la largeur de la personne détectée ou l'historique de données de santé du sujet permet une amélioration significative des

performances de leur application, engendrant un taux de bonne détection des chutes de $79{,}8\%$ à l'aide de ces informations, contre 68% sans leur utilisation.

En 2007, Lin et Ling [51] proposent une méthode de détection des chutes dans le domaine compressé (images MV et DC+2AC). Leur système utilise un serveur local collectant les données issues de caméras vidéo, compressées en MPEG-4. Les données sont ensuite envoyées via internet au centre de traitement et de contrôle. Leur algorithme est décomposé en deux étapes : (i) l'extraction de l'objet dans le domaine compressé ; (ii) la détection de la chute. L'extraction est réalisée à l'aide d'une détection globale de mouvement, permettant d'obtenir le mouvement global et le mouvement instantané. L'ensemble est ensuite combiné pour obtenir un masque de l'objet d'intérêt, affiné par la suite à l'aide d'un module de détection des modifications. L'étape de détection de la chute, quant à elle, utilise trois paramètres : (i) le barycentre de l'humain extrait ; (ii) la projection verticale de l'histogramme de l'objet ; (iii) la durée de l'événement. Leur expérimentation utilise 78 séquences (48 séquences d'entraînement et 30 séquences de test).

Une approximation elliptique de la personne a été introduite par Foroughi et al. en 2008 [52, 53, 54] pour la classification des chutes en trois événements : (i) activités de la vie quotidienne ; (ii) comportement anormal ; (iii) chute, lui-même séparé en trois classes : chutes vers l'avant, vers l'arrière et latérales. La première étape de leur système consiste en une extraction de la silhouette de la personne à l'aide d'une procédure d'extraction de fond. Une ellipse est ensuite appliquée, contentant la silhouette détectée. De cette silhouette sont extraits les écart-types de l'orientation et du ratio entre les axes mineurs et majeurs de l'ellipse. Les projections des histogrammes sont calculés et la position de la tête est estimée à partir du point le plus haut dans l'image de l'ellipse. Enfin la posture de la personne est classifiée à l'aide d'un réseau de neurones pour [52, 54] et un Support Vector Machine (SVM) pour [53]. Pour [54], l'espace des caractéristiques extraites est réduit à l'aide de l'application d'une Analyse en Composantes Principales (ACP).

Williams et al. [39] proposent l'utilisation d'un réseau distribué de caméras. Pour ce faire, ils utilisent plusieurs noeuds de caméras sur batterie, chacun étant composé d'une caméra de faible fréquence ($1/5Hz$), d'un processeur, d'une RAM, d'une mémoire flash et d'un système de communication sans-fil. Ces caméras sont positionnées dans la pièce de façon à obtenir un point de vue se recouvrant. Enfin, une de ces caméras seulement est calibrée. Leur algorithme peut être décomposé en trois étapes distinctes réalisées au niveau de chaque noeud simultanément : (i) la détection de la personne ; (ii) l'homographie de l'image avec les noeuds voisins ; (iii) la détection de chute et sa localisation. La détection de la personne est effectuée à l'aide d'une soustraction de fond. Une SVM est utilisée afin de détecter une chute pour classifier les événements à partir du ratio hauteur/largeur. L'homographie, quant à elle, est effectuée à l'aide d'une DLT (Direct Linear Transform). Finalement, la localisation de la chute est permise par le recouvrement des points de vue des caméras et la calibration de la caméra maître.

Le système proposé par Cucchiara et al. en 2007 [55] est également basé sur une vue multi-caméra avec recouvrement du point de vue de la scène. La détection de la personne est effectuée à l'aide d'une soustraction de fond adaptative pour la suppression de l'effet "fantôme", c'est à dire l'aura laissée par un objet après un laps de temps inactif. L'utilisation de deux caméras au moins, calibrées dans le même système de coordonnées, leur permet d'estimer la silhouette de la personne lors de présence d'occlusion. Finalement, la classification des postures pour la détection des chutes est réalisée par un modèle de Markov caché (HMM).

Anderson et al. [56] utilisent un système multi-caméra pour réaliser un modèle 3D, appelé "personne voxel", de la personne à partir de la silhouette obtenue sur chaque caméra par soustraction de fond. Ils utilisent ensuite la logique floue pour mesurer l'état de l'objet à partir de trois états possibles : debout, sur le sol ou entre les deux situations.

Rougier et al. [57] proposent une méthode basée sur une analyse de la déformation de la forme de la personne à partir d'une correspondance de formes (shape matching). De l'image issue des caméras grand-angle est segmentée la silhouette de la personne. Une étape d'extraction de points à l'aide d'un détecteur de Canny est ajoutée afin d'effectuer une correspondance entre deux silhouettes consécutives. Finalement, un modèle de mixtures de gaussiennes est créé pour analyser l'activité de la personne et détecter les chutes à partir de deux caractéristiques, l'une représentant la chute et l'autre l'absence de mouvement après la chute.

Une analyse des comportements humains à l'aide d'une classification de postures a été réalisée en 2005 par Cucchiara et al. [58]. Pour ce faire, ils proposent un système temps réel basé sur une architecture client/serveur. Leur méthode se décompose en deux étapes distinctes : (i) pour chaque image, la posture de la personne est extraite par projection d'histogrammes. Celle-ci est ensuite comparée à l'aide d'un classifieur bayésien avec des cartes de projections réalisées précédemment pour chaque posture lors d'une phase d'entraînement ; (ii) la posture obtenue est ensuite analysée temporellement afin de vérifier la pertinence de la posture détectée par rapport aux informations de suivi. Leur méthode permet une bonne classification de 95% des postures.

Thome et al. [59] proposent une détection des chutes à l'aide d'une modélisation du mouvement utilisant un modèle de Markov caché hiérarchique (HHMM) à deux couches. La première couche utilise la posture de la personne et est appelée "motifs du comportement". Elle permet la description de variations brusques du mouvement. Celle-ci est divisée en deux états, correspondant à une personne debout et une personne à terre. Finalement, la seconde couche, utilisant comme états la première couche, permet la description du mouvement global comme une suite de mouvements élémentaires de la première couche. Leur système présente l'avantage d'être à même de décrire des mouvements complexes à l'aide de leur décomposition en suite de mouvements simples.

1.3.3 Systèmes basés sur le suivi vidéo de la tête

Afin de prendre en compte l'ensemble des phases de la chute décrites précédemment, un certain nombre de méthodes utilisent l'information de la position 3D de la tête pour la détection de la chute. La richesse spatio-temporelle de cette approche permet d'affiner l'analyse et de l'adapter à d'autres applications comme la reconnaissance des activités quotidiennes.

En 2005, Rougier et Meunier [60] développent un système de détection des chutes basé sur la vitesse 3D de la tête de la personne suivie. Pour ce faire, ils proposent le suivi 3D de la tête à partir d'une unique caméra. Leur méthode se décompose en trois étapes : (i) le suivi de la tête en deux dimensions ; (ii) le suivi 3D à l'aide d'un filtre particulaire ; (iii) la détection de chute à partir de la vitesse de la tête. L'information de profondeur à partir d'une image 2D est obtenue en utilisant l'algorithme de Dementhon et Davis [61] : la tête est représentée par une ellipsoïde 3D projetée dans le plan 2D. L'algorithme prend trois paramètres : les points 3D du modèle, les points correspondants en 2D dans le plan image et les paramètres intrinsèques de la caméra vidéo. Enfin, leur système utilise un ensemble de caméras IP reliées par un routeur à une unité de traitement, permettant l'envoi d'un signal d'alerte à un centre d'urgence [62].

Hazelhoff et al. [63] proposent en 2008 l'utilisation de deux caméras fixées perpendiculairement et non-calibrées pour la connaissance de l'information de la position 3D de la personne et la détection des chutes à l'aide de la vitesse 3D. Leur système peut se décomposer en cinq étapes : (i) la segmentation de l'objet avec une soustraction de fond sur chacune des deux caméras ; (ii) le suivi de l'objet détecté sur les deux caméras ; (iii) l'extraction de la direction de l'axe principal du corps et le rapport des variances des directions x et y à l'aide d'une ACP ; (iv) l'application d'un classifieur gaussien pour la reconnaissance des chutes ; (v) le suivi de la tête sur les images précédant la chute pour réduire le nombre de fausses alarmes. Leur système permet un taux de bonne détection de l'ordre de 85%.

Finalement, Auvinet et al. [46] utilisent un réseau de caméras partageant un large point de vue et calibrées pour l'estimation de la forme et de la position de la personne. La prise en compte de la hauteur de la tête de la personne est prise en compte pour la détection de la chute. Leur méthode, expérimentée sur 14 scénarios de chute et 14 activités normales de la vie quotidienne, atteint un taux de 100% de bonne détection. Néanmoins, leurs résultats sont à nuancer par la relative faible étendue de leur jeu de données. Une étude de leur système suivant le nombre de caméras utilisées (de 3 à 5) a été effectué en 2011 [64]. Ils atteignent $80{,}6\%$ de bonne détection pour 3 caméras contre près de 100% pour 5 caméras. Cependant, leur base de données de test ne comprend qu'une seule personne.

1.3.4 Discussion

Ainsi, trois approches majeures se distinguent pour la détection des chutes dans l'habitat à l'aide de l'information visuelle. Ces trois approches ne se focalisent pas sur les mêmes phases de la chute.

Les approches basées sur l'inactivité disposent de l'avantage de la simplicité du critère de détection de chute. En revanche, elles ne permettent pas l'analyse complète de la situation. Le système proposé par Jansen et Deklerck [48] s'affranchit de cette limitation en combinant l'information de l'orientation de la personne avec le critère d'inactivité.

La plupart des systèmes basés sur la détection de la posture de la personne utilisent une soustraction de fond pour la segmentation de l'individu. Cette opération, peu coûteuse en calcul, permet d'extraire rapidement le squelette de la personne et finalement de déterminer un ratio hauteur/largeur afin de connaître l'orientation de la personne. Le désavantage de ces méthodes est, tout comme les approches basées sur l'inactivité, qu'elles ne permettent généralement de détecter qu'une unique phase de la chute, la phase "post-chute". Néanmoins, cette approche ne génère que peu de fausses alarmes. De plus l'utilisation d'un rapport pour la détection de l'orientation est insensible à la distance de la personne par rapport à la caméra et ne nécessite donc pas la mise en place de systèmes d'estimation de la profondeur (e.g. stéréovision, caméras-TOF). Le fait qu'elles détectent la personne une fois à terre permet l'utilisation de systèmes de traitement des données minimalistes, réduisant d'autant les coûts d'utilisation. Enfin, elles bénéficient d'une grande pertinence quant à la réduction des fausses alarmes lors de leur utilisation en complément d'une seconde méthode de détection de chutes [12].

Finalement, l'utilisation de la trace en trois dimensions de la tête de la personne permet l'analyse de la chute dans sa globalité. Cet avantage majeur rend possible la différenciation des chutes molles et brutales, engendrant une information complémentaire utilisable par le personnel de santé. De plus, le suivi en temps réel de la personne est l'approche possédant les perspectives à long terme les plus prometteuses, rendant notamment possible l'apprentissage des activités de la vie quotidienne de la personne et donc l'adaptabilité du système, notamment pour la détermination de zones d'inactivité. De plus, cela permet une utilisation dans un système d'aide à la personne plus général pour une application à la détection précoce de maladies neuro-dégénératives. Malheureusement, l'obtention de l'information 3D nécessite une étape de calibration des caméras, réduisant l'adaptabilité du système.

La plupart des systèmes décrits utilisent des caméras standard (e.g. webcams, caméras IP). Cela est motivé par le faible coût de tels dispositifs. Cependant, quelques méthodes sont basées sur des caméras grand-angle afin de pouvoir obtenir sur une unique image l'ensemble de la pièce. La grande distorsion des images et leur faible pertinence pour une utilisation de la position 3D de la tête de la personne constituent les principaux désavantages de ce type de capteurs.

1.3.5 Conclusion

Les approches basées sur l'information visuelle permettent une détection des chutes en s'affranchissant de capteurs portés par la personne et pouvant donc être oubliés par l'habitant. De plus, la grande richesse des informations visuelles rendent possible une application utilisant un unique type de capteur environnemental bien que le majeur problème des systèmes basés sur la vision est posé par la possibilité d'occlusion partielle ou totale de la personne. Trois approches de détection par vidéo-surveillance peuvent être observées. Bien que le suivi en trois dimensions représente la méthode la plus prometteuse, celle-ci souffre d'une plus faible robustesse, pouvant être contrebalancée par une combinaison des trois méthodes visuelles de détection de chutes. Dans cette thèse, nous nous orientons vers une méthode basée sur le suivi vidéo de la tête. Pour ce faire, il est nécessaire d'être en mesure de détecter l'objet nous intéressant dans une séquence vidéo.

1.4 Détection de l'objet dans une image

La mise en place d'un système de vidéo-surveillance pour la détection des chutes nécessite d'être à même de détecter la personne dans l'image. La détection et le suivi d'un objet dans l'image est un domaine en expansion grâce notamment à la performance grandissante des ordinateurs et de la qualité et le faible coût de caméras. Le problème posé peut-être décomposé en deux étapes majeures [65] : (i) le modèle de représentation de l'objet ; (ii) l'extraction de l'objet dans l'image. Évidemment, le choix de la représentation est dépendant de la méthode d'extraction utilisée. Plusieurs problèmes se posent, tels que la perte d'information causée par le passage d'un monde en trois dimensions à une image en deux dimensions, la sensibilité au bruit présent sur l'image, les occlusions partielles ou totales, les variation de la luminosité. On dénombre quatre grandes famille de détection d'objets : (i) la détection de points d'intérêts ; (ii) la soustraction de fond ; (iii) la corrélation ; (iv) l'apprentissage supervisé.

1.4.1 Détection de points d'intérêt

La détection de point est utilisée pour sélectionner des points variant peu avec la luminosité et le point de vue de la caméra. De nombreux détecteurs de points ont été et sont encore développés. Parmi les plus utilisés, on peut citer l'opérateur de Moravec, le détecteur de Harris, le Kanade-Lucas-Tomasi (KLT) ou le Scale Invariant Feature Transform (SIFT).

Opérateur de Moravec [66] Cette méthode se décompose comme suit :

1. calcul de la variation de l'intensité dans un bloc de taille 4×4 dans les directions horizontales, verticales, diagonales et antédiagonales ;
2. la valeur représentative pour le patch 4×4 correspond à la valeur minimale des 4 variations (directions) ;
3. un point est finalement sélectionné si sa variation d'intensité est un maximum local dans un bloc de 12×12.

Détecteur de Harris [67] Le détecteur de Harris est construit comme suit :

1. calcul des dérivées de premier ordre de l'image (I_x, I_y) pour déterminer les variations directionnelles d'intensité ;
2. application d'un moment d'ordre 2 (Eq. 1.1) autour de chaque pixel afin d'encoder cette variation d'intensité :
$$M = \begin{pmatrix} \Sigma I_x^2 & \Sigma I_x I_y \\ \Sigma I_x I_y & \Sigma I_y^2 \end{pmatrix}; \tag{1.1}$$
3. le point d'intérêt est finalement sélectionné en seuillant les variations de l'intensité locale R (Eq. 1.2), calculé à partir du déterminant et de l'intensité de M :
$$R = \det(M) - k.\,\mathrm{tr}(M)^2,\ k \text{ une constante.} \tag{1.2}$$

Kanade-Lucas-Tomasi (KLT) [68] L'algorithme KLT diffère du détecteur de Harris au niveau de la dernière étape. En effet, ici R est défini comme étant la valeur propre minimale de M (Eq. 1.3) :

$$R = \lambda_{min}. \tag{1.3}$$

Finalement, les points spatialement proches l'un de l'autre sont supprimés. Les algorithmes de Harris et KLT sont très proches et engendrent des résultats semblables.

Scale Invariant Feature Transform (SIFT) [69] Cette méthode, introduite pour remédier au principal défaut des détecteurs de Harris et KLT (leur sensibilité aux transformations projectives et affines), se décompose en quatre étapes :

1. L'image est convoluée avec des filtres gaussiens de différentes échelles afin de générer des images de différences de gaussiennes (DoG). Les minima et maxima locaux sont finalement extraits sur les images DoG à chaque échelle.
2. Interpolation des valeurs des couleurs.
3. Les points d'intérêts le long des bords ou de faible contraste sont éliminés.
4. Des orientations sont assignées aux points d'intérêt (pics de l'histogramme des directions de gradient).

Du fait que les points d'intérêt sont calculés à différentes échelles et résolutions, un nombre largement plus élevé de points d'intérêt est calculé que pour les détecteurs précédents, induisant une méthode beaucoup plus efficace [70].

1.4.2 Soustraction de fond

Le principe de la soustraction de fond est de séparer l'image en deux couches : l'arrière-plan, immobile, et le premier plan, en mouvement. Pour ce faire, il est impératif de créer une représentation du fond de la scène, et d'analyser les déviations de ce modèle pour chaque nouvelle image. Les changements significatifs correspondent finalement à un objet en mouvement.

L'approche proposée par [71] consiste à :

1. modéliser la couleur (domaine YUV) à l'aide d'une gaussienne 3D (les variables $\mu(x,y)$ et$\Sigma(x,y)$ sont calculées à partir de plusieurs images consécutives),

$$I(x,y) \sim \mathcal{N}(\mu(x,y),\Sigma(x,y)) \; ; \tag{1.4}$$

2. la ressemblance entre chaque pixel d'une image entrante et le fond est calculée ;
3. les pixels qui diffèrent sont considérés comme faisant partie d'un objet en mouvement.

Une amélioration de ce modèle utilisant un modèle statistique multimodal a été effectuée par [72]. Cette méthode utilise une mixture de gaussiennes pour modéliser la couleur du pixel. Le principe est donc :

1. création d'un modèle pour le fond de l'image ;
2. comparaison d'un pixel avec l'ensemble des gaussiennes du modèle ;
3. si une correspondance est trouvée, la moyenne et la variance de la gaussienne est mise à jour, si aucune correspondance n'est trouvée, une nouvelle gaussienne est ajoutée à la mixture.

Il existe certaines méthodes utilisant des informations sur la texture, ayant l'avantage d'être peu sensible aux changements de luminosité. L'utilisation des modèles de Markov cachés (HMM) est également avantageuse pour certains évènements difficiles à modéliser avec les approches précédentes [65].

L'approche par la décomposition en sous-espaces propres, proposée par [73] permet également de s'affranchir des problèmes engendrés par la variation de la luminance. Le fond est modélisé grâce à la décomposition et l'image d'entrée est ensuite projetée sur la matrice de vecteurs propres, permettant de détecter le premier plan (l'objet en mouvement).

Les avantages de la soustraction de fond sont ses capacités à modeliser les changement de la luminance, le bruit et à détecter les objets en déplacement. Par contre, cette méthode peut détecter des régions incomplètes, et est très sensible aux déplacements de la caméra. En outre, un effet de fantome est susceptible d'apparaître lorsqu'un objet auparavant immobile se met subitement en mouvement.

1.4.3 Corrélation

Issue du domaine de l'optique, la corrélation a profité de l'apparition d'unités de traitement numériques. Elle prend ses fondements dans la comparaison d'images dans le plan de Fourier. Le schéma général peut se décomposer comme suit :

1. sélection d'une image de référence de l'objet à reconnaître ;
2. calcul de la transformée de Fourier de l'image de référence ;
3. fabrication du filtre de corrélation ;
4. calcul de la transformée de Fourier de l'image cible (e.g. image courante de la séquence vidéo) ;
5. multiplication du spectre de l'image cible avec le filtre de corrélation ;
6. transformée de Fourier inverse du module du plan obtenu à l'étape précédente ;
7. récupération du plan de corrélation, présentant un maximum, appelé "pic de corrélation", donc l'intensité permet d'évaluer le niveau de ressemblance entre l'image cible et l'image référence.

Cette méthode, qui sera explicitée en détail dans le chapitre 2, permet à la fois la détection, la localisation et l'identification.

1.4.4 Segmentation

Le but de la segmentation est de partitioner l'image en des régions distinctes. On peut par exemple dénombrer les méthodes de Mean-Shift Clustering et les contours actifs.

Mean-Shift Clustering [74] Le principe de la Mean-Shift Segmentation est de trouver des classes dans l'espace espace+couleur $[l,u,v,x,y]$, où $[l,u,v]$ représente la couleur et $[x,y]$ la localisation spatiale. Le principe est énuméré comme suit :

1. l'algorithme est initialisé en positionant au hasard des classes (ellipsoïdes multidimensionnels) hypothétiques ;
2. chaque centre de classe est déplacé à la moyenne des données à l'intérieur de la classe, créant ainsi un "mean-shift vector" entre l'ancien et le nouveau centre de la nouvelle classe ;
3. cette technique est ensuite recalculée de manière itérative jusqu'à ce que les centres des classes restent immobiles.

Cette methode est très sensibles aux paramètres choisis, telle que la taille minimale de chaque classe.

Contours actifs [75] Le principe de cette méthode est de faire évoluer un contour autour des limites de l'objet suivi. L'évolution du contour est faite à l'aide d'une fonction de l'énergie, définissant la proximité du coutour avec la région de l'objet :

$$E(\Gamma) = \int_0^1 (E_{int}(v) + E_{im}(v) + E_{ext}(v))ds \tag{1.5}$$

où s est la longueur du contour Γ, E_{int} inclut des contraintes de régularisation, E_{im} inclut l'énergie basée sur l'apparence, E_{ext} inclut des contraintes additionnelles. La méthode des contours nécessite une initialisation, demandant une connaissance à priori soit du fond, soit de l'objet. De plus, elle nécessite un traitement lourd en calculs, la rendant peu adaptée au temps réel

1.4.5 Apprentissage supervisé

L'apprentissage supervisé permet d'apprendre différentes vues de l'objet de façon automatique à partir de données d'entrainement. La sortie d'un tel système peut se faire sous la forme d'une valeur continue, dans ce cas le système est appelé "régresseur", ou d'un label de classe, le système est donc un "classifieur". Différentes méthodes existent, telles que les réseaux de neurones, les arbres de décision ou encore le boosting adaptatif et le SVM (Support Vector Machine).

Boosting adaptatif Le principe du boosting est de construire itérativement un classifieur à partir de plusieurs classifieurs basiques [76]. Cette méthode consiste à :

1. construire une distribution de poinds à partir du training set ;
2. sélectionner un classifieur retournant la plus faible erreur (proportionnelle au poids des données mal classées) ;
3. le poids des données mal classées est augmenté.

Dans cette méthode, l'ensemble des classifieurs de base peut être considéré comme un ensemble de seuils.

SVM (Support Vector Machines) L'approche SVM permet de séparer les données en deux classes (objet ou non objet). Cela est effectué en recherchant le meilleur hyperplan permettant de maximiser la marge, c'est-à-dire la distance entre l'hyperplan et les données les plus proches, les "vecteurs de support" [77].

1.4.6 Conclusion

Les méthodes basées sur la détection de points d'intérêt permettent un suivi rapide et robuste de l'objet. Elles sont généralement adaptées à des objets représentant une petite portion de l'image, ce qui n'est pas le cas dans un système de détection de chutes. Cependant, il est possible d'utiliser un grand nombre de points d'intérêt pour représenter un objet. La soustraction de fond quant-à-elle est particulièrement adaptée à la détection de postures. En effet, une simple étape d'érosion permet une extraction du squelette. Pour ce qui est de la segmentation, il s'agit d'une méthode très efficace mais souffrant de son coût en calcul important. De plus, elle nécessite une initialisation manuelle. La corrélation est une méthode rapide et ayant le principal avantage d'être adaptée à la détection, la localisation et l'identification de l'objet dans une image. Pour cette raison, nous proposons l'application de la corrélation pour la détection et l'identification de la personne dans le cadre de notre système de détection de chutes.

1.5 Notre système de détection de chutes

Notre système fait suite à une nécessité, formulée par la mutuelle de santé Malakoff-Médérik, de réalisation d'une solution de détection de chute à faible coût de fabrication et d'installation et s'affranchissant totalement du port de capteurs par la personne dépendante. Par conséquent, il est nécessaire d'imaginer une solution capable de détecter et d'analyser une situation dans le but d'interpréter une situation. Il est de plus primordial que ce système ne repose pas uniquement sur des capteurs portés, ceux-ci pouvant apporter une précision supérieure ou des informations complémentaires sur l'état de santé de la personne (e.g. constantes physiologiques) mais ne devant pas être indispensables à la détection d'une chute ou d'un signal de détresse. Un tel système doit également être opérationnel continuellement sans intervention de la part de l'utilisateur, que ce soit de jour, laps de temps pendant lequel les conditions de luminosité sont optimales pour un système basé sur des caméras vidéo, ou de nuit.

Le système doit être déployable le plus largement possible, aussi bien dans de futurs centres d'hébergement pour personnes âgées ou dépendantes, que dans des centres déjà existants, des hôpitaux ou directement

dans l'habitation de la personne. En effet, aussi bien pour des raisons pratiques (besoin de diversification et d'augmentation des capacités de prise en charge des personnes dépendantes) que sanitaires (apparition tardive et ralentissement de la progression des maladies neuro-dégénératives) ou sociales (préservation du lien social), il est nécessaire de trouver des solutions pour permettre une vie à domicile des personnes faiblement dépendantes. Dans ce but, la contrainte centrale est donc d'obtenir un système dont l'installation et la fabrication puissent se faire à faible coût, sans nécessiter d'importants travaux dans l'habitation. Ainsi sont écartés des systèmes basés sur un sol sensible à la pression, tels que le Smart Floor [34].

Également, la solution mise en place doit être capable de fonctionner en temps réel, c'est-à-dire de traiter les données acquises au fur-et-à-mesure de leur acquisition (e.g. fréquence de rafraîchissement de la caméra vidéo). En outre, une fois une chute détectée, les éventuels algorithmes de vérification supplémentaires doivent permettre l'envoi d'une alarme à un personnel de soin dans un laps de temps acceptable pour la survie de la personne (quelques secondes).

La méthode doit être à même d'identifier la personne présente dans la pièce. En effet, un tel système peut être amené à être utilisé dans un lieu où vivent plusieurs personnes (e.g. centres de soins et d'hébergement), la personne dépendante peut recevoir de la visite, notamment d'enfants, peu enclins à provoquer une alarme manuellement si une situation d'urgence se présente. Ainsi, l'identification est nécessaire pour déterminer la pertinence de suivre le visage qui vient d'être détecté. Cela évite une focalisation de l'algorithme sur une personne non à risque et permet donc d'alléger l'occupation du processeur. En outre, cela permet aussi d'écarter une détection un décèlement d'un visage dans l'image alors qu'il n'y en a pas en réalité, voire d'un visage présent sur un tableau dans la pièce. De plus, la reconnaissance de la personne suivie permet d'enregistrer certaines données liées aux habitudes de la personne, pouvant être intéressantes pour un praticien, notamment pour diagnostiquer précocement des maladies neuro-dégénératives [78]. Enfin certains paramètres peuvent être adaptés à la personne suivie, comme la taille de la personne et donc le critère de détection des chutes ou la taille de la région de recherche du visage.

Pour répondre à cela, nous proposons un premier système de vidéo-surveillance basé sur des caméras basse résolution, de type webcam. Celui-ci pourra être complété par d'autres capteurs tels que des caméras infrarouge pour la vision de nuit ou des détecteurs de porte ou de mouvement. Un aperçu de notre système est présenté en figure 1.4. Moyennant des travaux de recherche supplémentaires, de nouveaux capteurs pourront être implantés, c'est-à-dire des microphones, capteurs infrarouge ou de pression. Une unité de calcul locale est chargée de la fusion des différentes données. Ainsi, il sera donc possible de considérer une large variété de situations, telles que les chutes, une absence d'action, ou un changement soudain dans les habitudes de la personne, rendant possible accession à des situations d'urgences mais également de fournir au praticien des informations facilitant le diagnostic. Finalement, un niveau d'alerte suivant le degré d'urgence pourra être émis aux proches parents, voisins ou à un centre d'intervention pour une réponse rapide.

Notre système de détection des chutes de la personne âgée doit donc être à même de traiter toute la chaîne de traitement, de l'acquisition et du traitement local de données jusqu'à l'envoi, si besoin, d'une alarme. Nous proposons donc de le décomposer en quatre parties distinctes : (i) un module d'identification ; (ii) un module de suivi ; (iii) un module de prise de décision ; (iv) un module de communication.

Parmi les méthodes de traitement d'images pour la détection et l'identification, nous nous sommes focalisés sur la corrélation. En effet, celle-ci, bien qu'étant peu employée dans des applications récentes, présente l'avantage d'être capable d'analyser une image dans sa globalité. En outre, elle permet simultanément l'identification et la détection. Finalement, s'agissant de la compétence de notre laboratoire, il était impératif d'explorer ses capacités dans une application concrète de suivi d'une personne.

Les chapitres 2 à 4 traitent de l'utilisation de la corrélation et de leur application pour l'identification et le suivi d'objets. Le chapitre 2 consiste en une présentation de la corrélation et des deux principales architectures, le corrélateur à spectre joint (JTC) et le corrélateur de Vander Lugt (VLC) dans le but de déterminer leur pertinence dans les deux applications visées. L'identification par corrélation, le premier module de notre système est présenté dans le chapitre 3. Dans ce chapitre, nous proposons l'utilisation du corrélateur de Vander Lugt pour l'identification et nous proposons une décomposition en modèle linéaire du plan de corrélation pour

l'extraction du pic de corrélation. Le chapitre 4, quant à lui, est consacré au suivi d'un objet par corrélation. La localisation pour le suivi est effectuée à l'aide de l'architecture JTC. Notre algorithme de suivi est de type itératif : la région détectée dans l'image cible est utilisée comme référence pour l'identification suivante. Enfin, l'ensemble du système, utilisant les algorithmes d'identification et de suivi par corrélation, présentés au chapitres 3 et 4, est développé dans le chapitre 5.

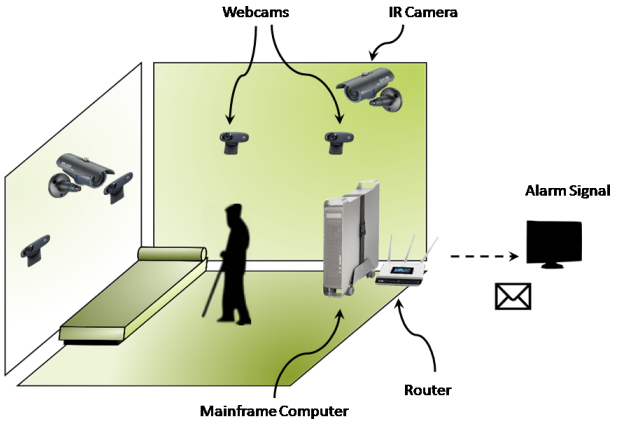


FIGURE 1.4 – Aperçu de notre système de détection des chutes.

Chapitre 2

La corrélation optique

Sommaire

Durant les deux dernières décennies, de nombreux travaux ont été effectués dans le domaine de la reconnaissance des formes [79, 80]. Ceci est dû au potentiel de son application dans des domaines aussi variés que la sécurité ou le paramédical. De nombreuses approches se sont dégagées, pouvant être basées sur la corrélation optique, comme le corrélateur de Vander Lugt (VLC) ou le corrélateur à spectre joint (JTC). On retrouve également des méthodes numériques comme les eigenfaces [81], les ondelettes [82], l'analyse en composantes principales (PCA) [83] ou l'analyse en composantes indépendantes (ICA) [84, 85]. Dans ce chapitre, nous nous intéressons à la corrélation optique et ses applications numériques, notamment dans le cadre de la détection et la reconnaissance d'objets.

La production scientifique sur la corrélation s'est essoufflée significativement durant les dernières années. Cela peut-être expliqué par deux raisons principales. Premièrement, beaucoup de travaux ont été menés sur la proposition et la validation de filtres de corrélation [86, 87, 88], au détriment des pré- et post-traitements. Deuxièmement, cela est dû à une concentration de la recherche sur une implantation tout-optique, compliquant grandement l'utilisation de telles méthodes. À l'heure actuelle, le développement d'unités de traitement performantes et peu coûteuses (GPU, FPGA [89]) permettent une utilisation de la corrélation à l'aide d'une implantation numérique, ouvrant la voie à de nouvelles optimisations de cette méthode prometteuse.

Dans ce chapitre, nous présentons tout d'abord le principe général de la reconnaissance par corrélation ainsi que les métriques utilisées pour quantifier la reconnaissance et mesurer les performances du classifieur, c'est-à-dire les capacités de séparation des résultats en deux classes distinctes. Ensuite, nous introduisons le corrélateur de Vander Lugt (VLC) et une étude de ses performances. Finalement, nous présentons le corrélateur à spectre joint (JTC) et quelques-unes de ses variantes.

2.1 Principes

Le principe de la corrélation consiste, de façon simpliste, à comparer une image cible (image à reconnaître) avec une image de référence (à partir d'une ou plusieurs images connues préalablement) afin d'en évaluer la ressemblance. La corrélation est exprimée mathématiquement par l'équation 2.1 :

$$(\overline{h} * i_{Cible})(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \overline{h}(x-s,y-t) i_{Cible}(x,y) ds dt. \tag{2.1}$$

où $\overline{h}(x,y)$ est le conjugué du filtre de corrélation $h(x,y)$, i_{Cible} est l'image cible et $*$ le produit de convolution.

Le théorème de Plancherel [90] donne l'expression de la corrélation en fonction du produit des transformées de Fourier TF de h et i_{Cible} :

$$(\overline{h} * i_{Cible})(x,y) = TF^{-1}[\overline{H}.I_{Cible}], \tag{2.2}$$

où $\overline{H}$ et I_{Cible} sont les transformées de Fourier de Fourier de h et i_{Cible}, respectivement, où et TF^{-1} correspond à la transformée de Fourier inverse. La transformée de Fourier est définie par l'expression (Eq. 2.3) :

$$TF[f] = \int_{-\infty}^{+\infty} f(x) e^{-2j\pi\nu x} dx, \tag{2.3}$$

où ν est la fréquence du signal.

La corrélation nous donne une mesure de ressemblance entre une image référence, à reconnaître, et une image cible, issue par exemple d'une base de données. Cette méthode effectue une analyse sur une région entière de l'image (contrairement aux approches locales basées sur des traits et points caractéristiques). L'intérêt de l'utilisation de la corrélation réside dans le fait qu'elle peut être réalisée expérimentalement à l'aide d'un montage optique. Cette caractéristique autorise un temps de calcul extrêmement rapide, celui de la vitesse de

propagation de la lumière dans l'air et le verre. Cependant, un tel montage est difficile et coûteux à mettre en oeuvre. La miniaturisation et le déploiement à grande échelle d'unités de calcul (CPU, GPU, FPGA) a rendu possible une implantation numérique de cette famille de méthodes [91]. Nos travaux reposent sur de telles implantations, permettant l'application d'algorithmes numérique de traitement des données, en pré- ou post-traitement.

Deux principales architectures optiques sont présentées dans la littérature : le corrélateur à spectre joint (JTC, "Joint Transform Correlator") et le corrélateur de Vander Lugt (VLC, "Vander Lugt Correlator"). Ces deux architectures sont basées sur l'implantation optique dite "4f", présentée en figure 2.1. Cette installation consiste à illuminer un plan d'entrée, contenant l'image cible pour le VLC et l'image cible et l'image référence pour le JTC. Celui-ci est placé sur le foyer en amont d'une lentille convergente. Sur le foyer placé en aval de cette lentille se trouve le plan de Fourier (le placement sur le foyer d'une lentille convergente d'une image permet d'obtenir sa transformée de Fourier sur le foyer situé après la lentille). À ce niveau et en ce qui concerne le VLC, le plan de Fourier est multiplié avec un filtre de corrélation à l'aide d'un "Spatial Light Modulator" (SLM). Pour le JTC, une opération non-linéaire est appliquée. Finalement, une seconde lentille positionnée de la même façon que la première nous permet de calculer la transformée de Fourier inverse et donc d'obtenir un plan de corrélation. C'est ce positionnement basé sur quatre distances focales (deux par lentille) qui a donné l'appellation dite "montage 4f".

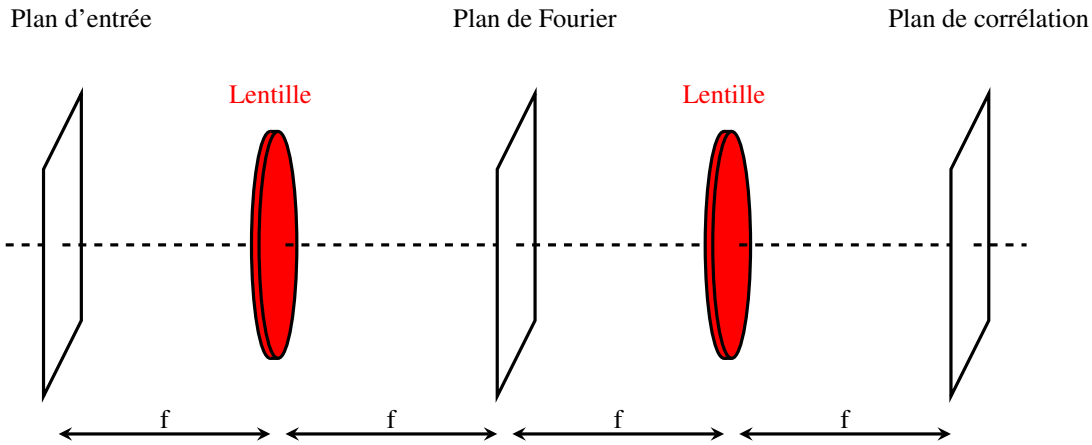


FIGURE 2.1 – Architecture 4f optique

Le plan de corrélation obtenu contient dans tous les cas au moins un pic de corrélation dans le cas du VLC et trois pics dans le cas du JTC (deux pics d'inter-corrélation et un pic d'auto-corrélation). Un exemple de plan de corrélation calculé à l'aide du corrélateur de Vander Lugt est présenté en figure 2.2. La figure 2.2a correspond à une corrélation entre une image référence et cible identiques tandis que la figure 2.2b correspond à une corrélation entre deux images issues d'objets différents. Nous observons un pic central de corrélation entouré d'un bruit de corrélation pour la corrélation vraie (Fig 2.2a) et du bruit de corrélation pour l'ensemble du plan pour la corrélation fausse (Fig. 2.2b). La hauteur du pic de corrélation est conditionnée par le degré de ressemblance entre l'image référence et l'image cible. Différents filtres de corrélation ont été développés : soit de façon optique [86, 92, 93, 94, 95, 96, 97], soit de façon numérique [98, 79]. Les filtres agissent sur l'image référence ou son spectre et permettent d'obtenir des plans de corrélation comprenant des pics de corrélation différents suivant la nature du filtre utilisé. Certains ont tendance à favoriser la robustesse, c'est-à-dire la capacité de reconnaissance du filtre lorsque l'image cible varie légèrement de l'image de référence, d'autres la discrimination, privilégiant une grande ressemblance entre l'image référence et cible. Le principal enjeu de la recherche utilisant des méthodes de corrélation restant, à ce jour, de déterminer le meilleur compromis

entre robustesse – la préférence est donnée à la reconnaissance de l'objet dans un maximum de situations (i.e. déformations, changement de point de vue...) – et discrimination – la préférence est donnée au rejet des fausses détections – pour l'application visée.

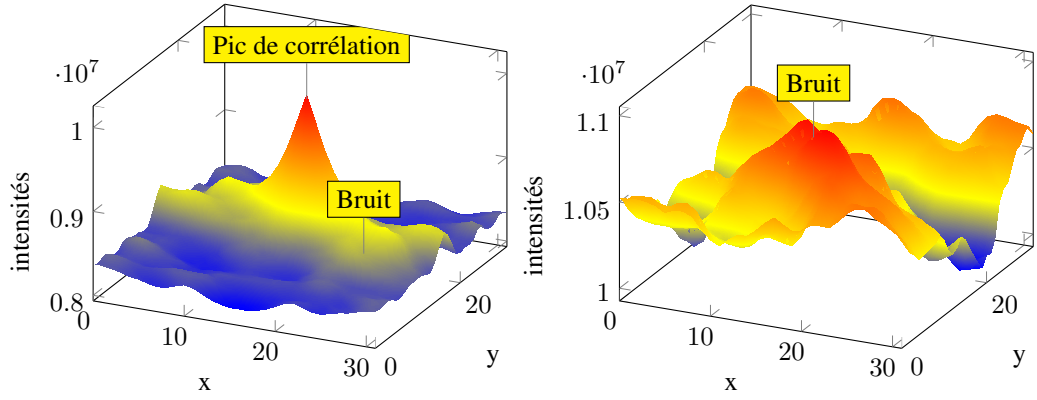


FIGURE 2.2 – Plan de corrélation du filtre adapté (cf. partie 2.4.1.1, page 35) : (a) les images cible et référence sont identiques ; (b) les images cible et référence sont issues de deux objets différents.

2.2 Critères d'évaluation de la corrélation

2.2.1 Critères de détection du pic de corrélation

Afin de pouvoir comparer les différents algorithmes de corrélation [95], il est nécessaire d'utiliser un critère objectif d'évaluation de la corrélation. Différents critères ont été proposés dans ce but, tels le rapport signal sur bruit (SNR), le peak-to-correlation energy (PCE) [99] et ses variantes, le PCE' et le PCE'' [100]. Tous ont pour point commun de calculer la prévalence du pic de corrélation, caractéristique de la corrélation, sur le reste du plan.

Le SNR, rapport des puissances du signal (pic de corrélation) et du bruit, et son équivalent en décibels, le SNR_{dB}, sont des mesures communément utilisées en traitement du signal. Le SNR est défini par l'équation (2.4) :

$$SNR = \frac{Puissance\ du\ signal}{Puissance\ du\ bruit} = \frac{|C(Pic)|^2}{\sqrt{\sum_{u=0}^{L}\sum_{v=0}^{H}|C(u,v)|^2}}, \tag{2.4}$$

où C_{Pic} est la valeur d'intensité du pic de corrélation (Pic étant les coordonnées du pic de corrélation) et $C(u,v)$ est l'intensité du pixel (u,v) du plan de corrélation,

$$\sum_{u=0}^{L}\sum_{v=0}^{H}C(u,v) = \sum_{u=0}^{L}\sum_{v=0}^{H}C_{Plan\ de\ corrélation}(u,v) - C(Pic) \tag{2.5}$$

correspondant au bruit, c'est-à-dire au plan de corrélation $\sum_{u=0}^{L}\sum_{v=0}^{H}C_{Plan\ de\ corrélation}(u,v)$ auquel on a retiré le pic de corrélation et L et H sont la largeur et la hauteur du plan de Fourier, respectivement.

Le SNR_{dB}, quant à lui, est exprimé par :

$$SNR_{dB} = 10\log_{10}(SNR). \tag{2.6}$$

Le PCE, est défini par l'énergie du pic de corrélation normalisée par l'énergie globale du plan de corrélation [100]. Il s'agit d'une modification par Kumar et Hasselbrook [101] de la métrique proposée par Dickey et Romero [99] :

$$PCE = \frac{\sum_{u=0}^{l}\sum_{v=0}^{h} E_{Pic}(u,v)}{\sum_{u=0}^{L}\sum_{v=0}^{H} E_{Plan\ de\ corrélation}(u,v)} = \frac{|C(Pic)|^2}{\sum_{u=0}^{L}\sum_{v=0}^{H} C_{Plan\ de\ corrélation}(u,v)^2}, \tag{2.7}$$

où E_{Pic} est l'énergie du pic de corrélation, $E_{Plan\ de\ correlation}$, celle du plan de corrélation. Pour un pic de corrélation intense, son énergie sera beaucoup plus élevée que celle du plan de corrélation et par conséquent le PCE calculé sera élevé, proche de 1. À l'opposé, un pic de corrélation étalé donnera un PCE avoisinant 0. Le plan de corrélation peu également contenir des pics périphériques, ne posant pas problème lorsque la corrélation est forte, mais pouvant engendrer des cas de fausse alarme lorsque le pic est peu intense.

Deux autres métriques, le PCE' et le PCE'' [100] ont également été proposées par Horner. Le PCE' a été réalisé afin de s'affranchir de la sensibilité du PCE et du SNR à la valeur moyenne du signal (à un niveau d'offset). Néanmoins, le PCE' est beaucoup plus sensible au bruit induit par le fond de l'image et par la finesse du pic de corrélation. À l'opposé, le PCE'' a été imaginé pour rendre compte des effets induits par le niveau d'offset et est par conséquent très influencé par celui-ci. Le PCE' est donc un compromis entre le SNR et le PCE et est défini par l'équation suivante :

$$PCE' = \frac{C(Pic)}{\sum_{u=0}^{L}\sum_{v=0}^{H} C(u,v)^2}. \tag{2.8}$$

Le PCE'' quant à lui est défini par :

$$PCE'' = \frac{PCE}{1 - PCE'} \tag{2.9}$$

Les performances de ces différents critères seront étudiées et comparées dans la partie 2.4.3.1 de ce chapitre.

2.2.2 Caractéristique de fonctionnement du récepteur

La caractéristique de fonctionnement du récepteur, ou Receiver Operating Characteristic (ROC) est une mesure des performances d'un classifieur. On la représente par une courbe paramétrée par un seuil de décision, présentant le taux de vrais positifs en fonction du taux de faux positifs.

La corrélation quant à elle permet de mesurer la similarité entre une image cible et une image référence. Il est donc possible d'effectuer une classification des résultats : leur séparation en deux groupes distincts suivant un seuil (le sujet est soit reconnu comme étant celui utilisé pour la création du filtre, soit non reconnu). Il s'agit donc d'un classifieur binaire. On définit le vecteur $\underline{w}$, appelé vecteur d'observation, composé de n observations $w_1, \cdots w_n$. Le principe est de prendre la meilleure décision à partir d'un ensemble d'observations suivant un critère $\hat{\theta}(\underline{w})$ qui soit la meilleure estimation du paramètre θ. On pose $\underline{Y}$ le vecteur formé par l'ensemble des valeurs prises par le critère d'évaluation lorsque le sujet cible correspond au sujet de référence, $\underline{Z}$ le vecteur formé dans le cas contraire. On définit l'espérance $E[\underline{Y}]$ de $\underline{Y}$, variable aléatoire de loi de probabilité discrète (p_i,y_i), contenant n éléments équiprobables par :

$$E[\underline{Y}] = \nu = \sum_{i=1}^{n} p_i y_i = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{2.10}$$

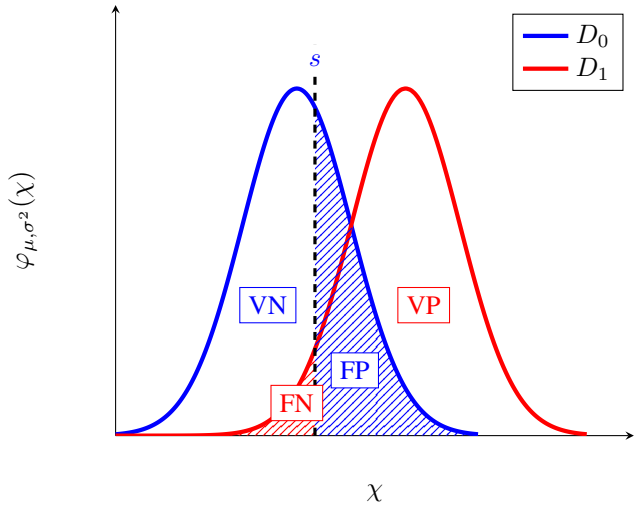


FIGURE 2.3 – Densités de probabilités $\varphi_{\mu,\sigma^2}(\chi)$ de la variable aléatoire χ séparée en deux classes D_0 et D_1 suivant un seuil s. VN correspond à Vrai Négatif, FN à Faux Négatif, FP à Faux Positif et VP à Vrai Positif.

La figure 2.3 illustre la séparation des observations $\underline{v}$ en deux classes indépendantes $\underline{Y}$ et $\underline{Z}$. Quatre différents cas sont possibles, les cas de détection et non détection, et les cas de fausse alarme et de non détection fausse, qui constituent des erreurs du classifieur, l'objectif étant de déterminer le seuil qui constituera le meilleur compromis. Afin de mesurer graphiquement la performance du classifieur, on utilise la courbe nommée "Receiver Operating Characteristic" (ROC) [102], courbe paramétrique représentant le taux de détection vrai TPR (appelé également sensitivité ou probabilité de détection vraie), en fonction du taux de fausse alarme FPR (appelé également "$1-$ spécificité" ou "probabilité de fausse alarme"), suivant la valeur du seuil de détection. Soit, en appelant H_0 l'hypothèse que l'image cible soit celle de la classe $\underline{Y}$ et H_1 que ce soit celle de la classe $\underline{Z}$, D_0 que la classe $\underline{Y}$ soit détectée et D_1 que la classe $\underline{Z}$ soit détectée, on a :

$$FPR = P(D_1|H_0) + P(D_O|H_1) \tag{2.11}$$

et

$$TPR = P(D_0|H_0) + P(D_1|H_1). \tag{2.12}$$

Nous considérons maintenant qu'une décision est prise en comparant l'estimateur $\hat{\theta}(\underline{w})$ avec un seuil s [102] :

$$\hat{\theta} \begin{cases} < s & \text{alors } D_0 \\ > s & \text{alors } D_1 \end{cases} \tag{2.13}$$

Les probabilités FPR et TPR s'expriment :

$$FPR = P(\hat{\theta}(\underline{w}) > s|H_0) + P(\hat{\theta}(\underline{w}) < s|H_1) \tag{2.14}$$

et

$$TPR = P(\hat{\theta}(\underline{w}) > s|H_1) + P(\hat{\theta}(\underline{w}) < s|H_0). \tag{2.15}$$

Pratiquement, la valeur du seuil est variée graduellement de 0 à 1. Selon la valeur de l'estimation $\hat{\theta}(\underline{w})$ et du seuil s, le classifieur prend une décision, D_0 ou D_1 , permettant la construction d'une matrice de confusion pour chaque valeur du seuil, comptabilisant le nombre de détections, de non-détections, de fausses alarmes et de non détections fausses (Tableau 2.1).

	Positif	Négatif
Positif	TP	FP
Négatif	FN	TN

TABLE 2.1 – Matrice de confusion

Les False Positive Rate (FPR) et True Positive Rate (TPR) sont finalement définis de la façon suivante :

$$FPR = \frac{FP}{FP+TN} \quad \text{et} \quad TPR = \frac{TP}{TP+FN}. \tag{2.16}$$

En traçant le FPR en fonction du TPR pour chaque valeur du seuil s, on obtient la courbe ROC paramétrée par le seuil s (Fig. 2.4).

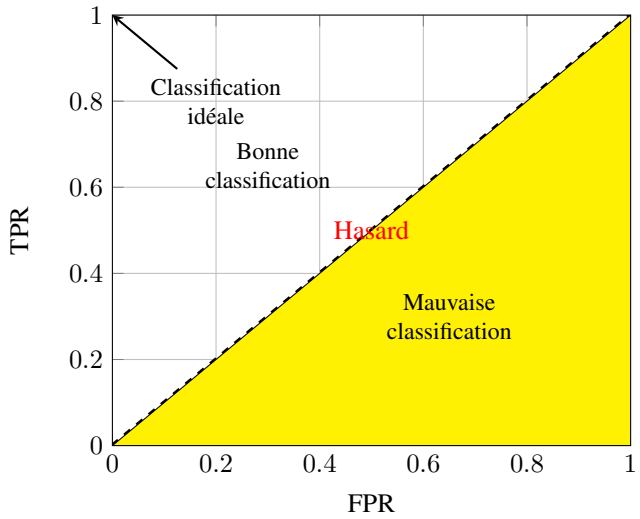


FIGURE 2.4 – Définition de la courbe ROC.

Le point dans le coin supérieur gauche du graphique, appelé "Perfect Classification", correspond au cas optimal, ou de détection singulière, apparaissant lorsque les densités de probabilité des vecteurs X et Y ont des supports disjoints. Lorsqu'un point de la courbe apparaît sur la diagonale, la classification est dite aléatoire. Cette diagonale divise donc l'espace de la courbe en deux parties : la bonne classification (au dessus de la diagonale) et la mauvaise classification (sous la diagonale). Une façon de représenter numériquement la performance d'un classifieur est donc de considérer l'Area Under Curve (AUC) [103]. On a donc, logiquement, $AUC = 0{,}5$ lors d'une classification aléatoire et $AUC > 0{,}5$ lors d'une bonne classification. Bien que cette métrique peut constituer une indication de la performance du classifieur, il s'agit d'une perte d'information par rapport à la courbe ROC elle-même.

2.3 Implantation optique ou numérique

Bien que l'implantation optique d'un corrélateur permette un calcul extrêmement rapide de la transformée de Fourier, ce temps de calcul est limité par les composants électroniques nécessaires. En effet, les différentes interfaces utilisées, le SLM permettant l'affichage du filtre ou du plan d'entrée, sont soumis à des limitations d'un autre ordre : fréquence de rafraîchissement du SLM et de la caméra CCD, capacité et temps d'accès des registres de stockage.

En outre, il existe d'autres inconvénients inhérents à un montage optique, à savoir sa complexité à implanter, notamment en ce qui concerne l'alignement des composants, et la présence d'aberrations chromatiques. Enfin, il s'agit d'une implantation relativement onéreuse.

Très récemment, l'arrivée sur le marché d'unités de calcul de plus en plus performantes et bon marché ont permis une implantation entièrement numérique de la corrélation optique. L'implantation numérique peut offrir notamment des capacités de souplesse d'utilisation, rendant possible la simulation de différents corrélateurs optiques sur un même composant suivant les besoins de l'application. La figure 2.5 résume la flexibilité de calcul des "systèmes sur une puce" (SoC) en fonction de leur rapidité. Ainsi, les composants les moins rapides, comme les CPUs (Central Processing Unit), seront ceux offrant la plus grande capacité de reprogrammation en fonction des besoins. Les composants les plus rapides,les circuits intégrés développés pour un client (ASIC), seront des unités développés uniquement pour une application donnée.

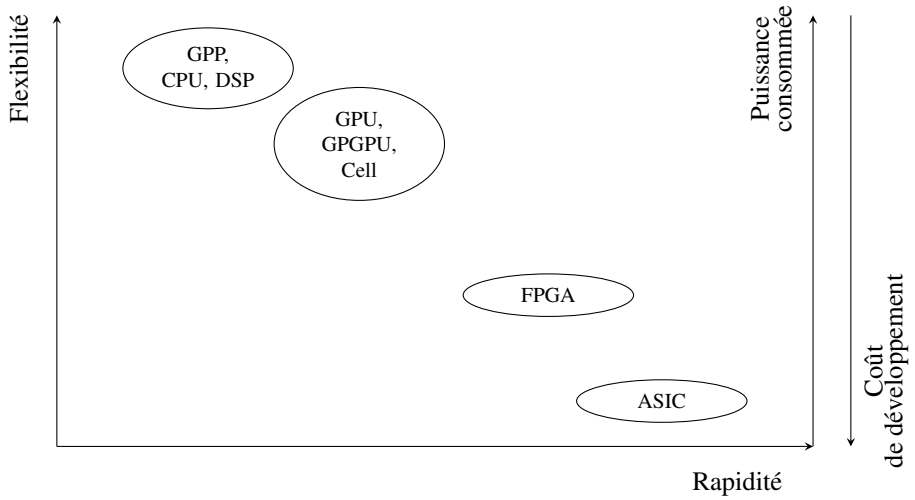


FIGURE 2.5 – Compromis Performance Flexibilité des SoC (tiré de [104]).

Les recherches présentées dans ce manuscrit portent principalement sur des applications nécessitant un certain nombre de traitements numérique de l'information contenue dans le plan de corrélation. En effet les travaux sont concentrés sur les capacités de suivi et de reconnaissance de la corrélation, et par conséquent titre nécessitent un traitement rapide des données. De plus, la nécessité d'un système implantable aisément nous oriente vers une implantation numérique de la corrélation.

2.4 Le Vander-Lugt Correlator

L'architecture du corrélateur de Vander Lugt est basée sur la multiplication du spectre de l'image cible (image à reconnaître) avec un filtre de corrélation H, réalisé à l'aide d'une image de référence. Cette approche

est illustrée dans le synoptique Fig. 2.6. La scène contenant l'objet à reconnaître, appelée "plan d'entrée", est multipliée avec le filtre de corrélation après application d'une transformée de Fourier. Le résultat, obtenu en effectuant la transformée de Fourier inverse, se présente comme un plan dit "de corrélation" présentant un pic central (proche d'un sinus cardinal en 2 dimensions), plus ou moins large et puissant suivant la ressemblance entre l'image cible et l'image de référence. Différents filtres ont été développés afin d'augmenter les performances de cette approche, initialement basée sur le Classical Matched Filter (CMF) [95], ou approche mono-corrélation – où la décision est prise sur une seule comparaison, ceci afin d'améliorer sa robustesse et sa discrimination. Une première tentative, largement décrite et explorée dans la littérature fut le Phase Only Filter (POF). D'autres méthodes ont finalement été introduites [86] afin de répondre au problème de la multi-corrélation, c'est-à-dire à la nécessité de comparer une image cible avec un set d'images de référence. Cela a notamment beaucoup été utilisé en reconnaissance de visages, dans le but de reconnaître une personne dans toutes les positions de l'espace et toutes les expressions du visage.

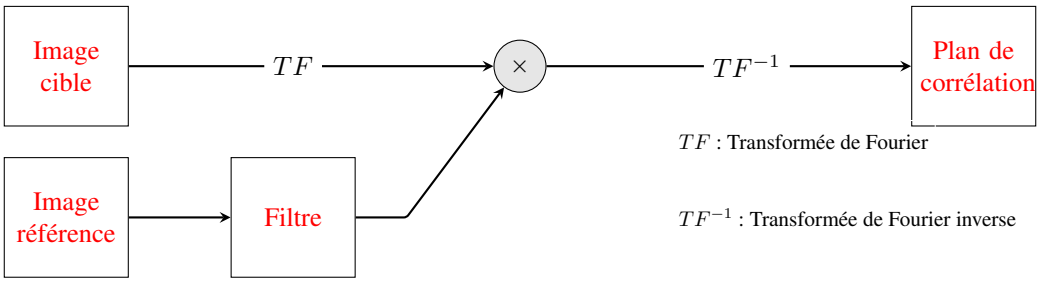


FIGURE 2.6 – Diagramme schématique du corrélateur de Vander Lugt.

2.4.1 Approche mono-corrélation

Afin de couvrir toutes les situations possibles de position ou d'expression d'un même objet, il est nécessaire d'inclure un nombre significatif d'images référence. Le principe de l'approche mono-corrélation est d'effectuer cette tâche au moyen d'un nombre de filtres égal au nombre d'image de référence. C'est-à-dire que pour chaque image de référence, l'algorithme nécessitera une étape de corrélation. Cette méthode permet une décision optimale mais avec un temps de calcul considérable.

2.4.1.1 Le filtre adapté

Le filtre adapté correspond à l'application directe de la théorie du signal. Le filtre est créé à partir du conjugué du spectre de l'image de référence :

$$H_{CMF}(u,v) = S^*_{R_1}(u,v) \tag{2.17}$$

où $S_{R_1}(u,v)$ est le spectre de l'image de référence R_1 et $S^*_{R_1}(u,v)$, son conjugué.

Cela équivaut donc à écrire :

$$H_{CMF}(u,v) = \rho_{R_1} e^{-i\phi_{R_1}} \tag{2.18}$$

où ρ_{R_1} est le module du spectre de l'image de référence, et ϕ_{R_1}, sa phase.

Afin de limiter le bruit de fond, il est possible de diviser le filtre ainsi obtenu par la densité spectrale de l'image. On obtient donc la définition suivante du filtre adapté :

$$H_{CMF}(u,v) = \frac{\alpha S^*_{R_1}(u,v)}{N(u,v)} \tag{2.19}$$

où $N(u,v)$ est la densité spectrale de l'image et α, une constante.

La figure 2.8 illustre une corrélation à l'aide du filtre adapté. Les images cible et référence sont présentées en figure 2.7. L'image référence utilisée est l'image Léna (Fig. 2.7a). Sont présentés un cas de bonne corrélation (Fig. 2.8a), à l'aide de Léna en image cible (Fig. 2.7a), et de mauvaise corrélation (Fig. 2.8b), où l'image cible (Fig. 2.7b) utilisée diffère de l'image référence. On obtient bien un pic de corrélation lorsque les deux images sont identiques et une absence de pic de corrélation bien visible lorsque les images sont différentes. Par contre, la taille est très large et donc peu discriminant et les PCEs ne diffèrent que très peu ($PCE = 0{,}0023$ pour des images identiques, contre $PCE = 0{,}0018$ pour deux images différentes). L'avantage de ce filtre est qu'il est très robuste mais il est malheureusement trop peu discriminant pour permettre un critère de reconnaissance efficace.

(a) (b)

FIGURE 2.7 – Images cible et référence : (a) image Léna ; (b) image Barbara.

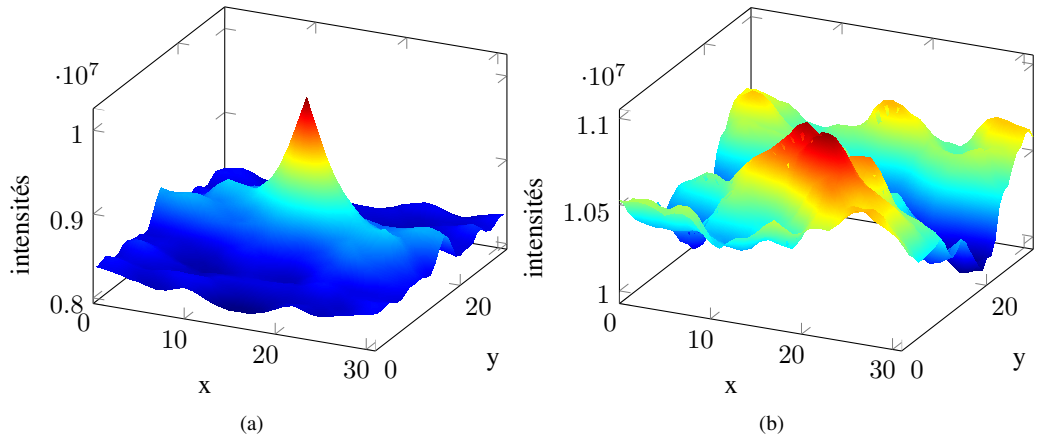
2.4.1.2 Le filtre de phase pure

Afin d'obtenir un filtre plus discriminant, le Phase Only Filter, conçu comme une optimisation du Classical Matched Filter, a été proposé [92]. Pour ce faire, on utilise la propriété suivante : la phase d'un spectre contient toutes les informations nécessaires pour reconstruire une image cible [105]. Ainsi on peut s'affranchir de l'amplitude du spectre qui présentera une très grande dynamique mais peu d'information. La figure 2.9 présente un exemple de cette propriété, la figure 2.9a est l'image originale et la figure 2.9b, l'image de phase. On observe que l'image de phase contient les informations de contours de l'image. Cette propriété permet d'obtenir des filtres beaucoup plus discriminants que le filtre adapté.

On remarque en effet que la phase permet de sélectionner uniquement les contours de l'image d'origine. Ce type de filtre est donc une version optimisée du filtre adapté, et se définit donc par l'équation :

$$H_{POF} = \frac{H_{CMF}}{\rho_{R_1}} \tag{2.20}$$

où ρ_{R_1} est le module du spectre,



(a) (b)

FIGURE 2.8 – Plan de corrélation du filtre adapté : (a) images référence et cible identiques ; (b) images référence et cible issues de deux personnes différentes

(a) (b)

FIGURE 2.9 – Sélectivité de la phase : (a) image d'origine ; (b) l'image de phase.

soit :

$$H_{POF} = e^{-i\phi_{R_1}} \tag{2.21}$$

d'où l'expression générale du filtre POF [1] :

$$H_{POF} = \frac{S^*_{R_1}(u,v)}{|S_{R_1}(u,v)|} \tag{2.22}$$

Un exemple d'application du filtre POF est présenté en figure 2.10. La figure 2.10a a été réalisée avec des images cible et référence identiques (Fig. 2.7a), tandis que la figure 2.10b l'a été avec des images différentes (Fig. 2.7a et 2.7b). On observe ici un pic de corrélation bien définit lorsque l'image testée est identique à l'image de référence, avec très peu de bruit de fond, et une absence de pic de corrélation lorsque l'image testée diffère de l'image de référence. De plus, le PCE est très élevé en présence du pic de corrélation ($PCE = 0{,}0869$) comparativement à la corrélation fausse ($PCE = 0{,}001$). Le filtre POF est donc un filtre extrêmement sélectif et donc bien plus discriminant que le filtre adapté, au détriment de la robustesse.

1. Phase Only Filter

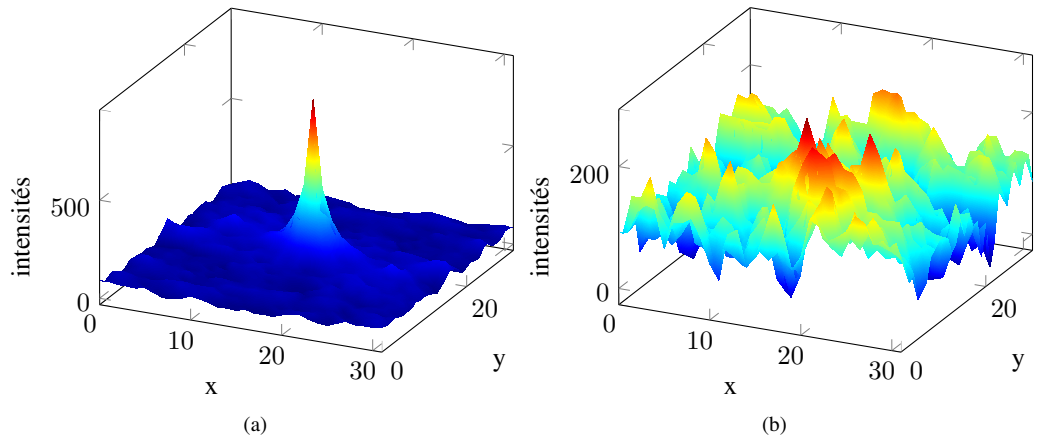


FIGURE 2.10 – Plan de corrélation du filtre de phase pure : (a) images référence et cible identiques ; (b) images référence et cible issues de deux personnes différentes

Un grand nombre de filtres dérivant des deux précédents ont été proposés dans la littérature. Nous retenons notamment les suivants :

- BPOF : filtre de phase pure binaire [106, 107],
- ROS_POF : filtre de phase pure avec région de support [107, 108],
- AMPOF : filtre de phase pure amplitude adaptée [93],
- CTPOF : filtre de phase pure complexe ternaire [109],
- FPF : filtre de puissance fractionnée [101],
- IF : filtre inverse [101],
- PCMF : filtre de phase à magnitude contrainte [110],
- PMF : filtre de phase principale [111],
- QPF : filtre à quadrature de phase [112, 113],
- OTF : filtre à compromis optimal [114].

L'ensemble de ces filtres ont été réalisés dans le but d'optimiser la décision prise à partir du pic de corrélation pour une seule image de référence. Les travaux présentés dans ce manuscrit ont été effectués afin d'améliorer les performances en travaillant directement sur le plan de corrélation. En ce sens, nous nous focalisons sur le filtre de phase pure. De plus, une optimisation considérable de la décision peut être effectuée en prenant en compte dans une même corrélation différentes images références.

2.4.2 Approche multi-corrélation

Afin de remédier aux problèmes de robustesse et de discrimination des précédents filtres, CMF et POF, et, en général, des filtres à une unique image de référence, l'approche multi-corrélation a été introduite. Le principe, schématisé en figure 2.11, est de composer un filtre à partir de plusieurs images référence, cela pour but de couvrir toutes les positions du visage, par exemple. Un avantage supplémentaire indéniable par rapport à une corrélation successive de plusieurs images de référence du même sujet est une réduction du nombre de corrélations et donc du temps de calcul, condition primordiale pour une implantation numérique qui est la situation se présentant dans nos travaux.

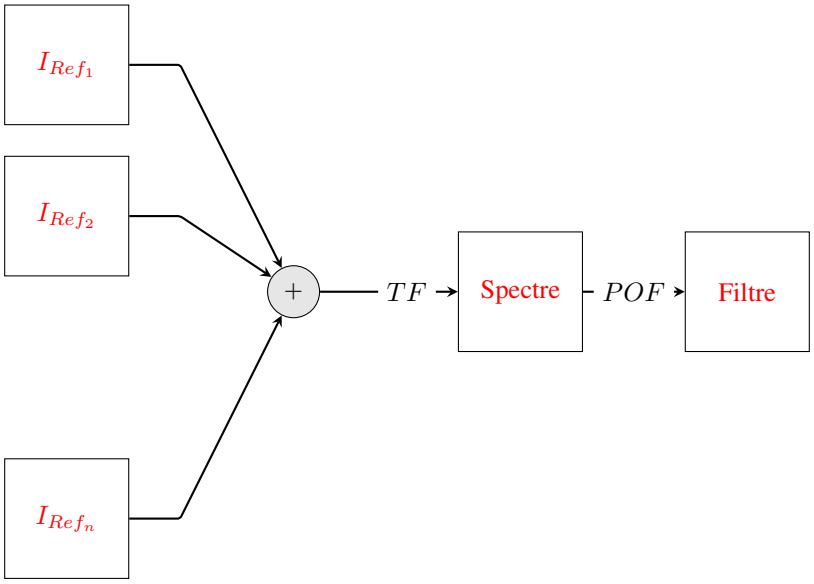


FIGURE 2.11 – Diagramme schématique du filtre composite.

2.4.2.1 Le filtre composite

Le filtre composite consiste en une combinaison linéaire des n différentes images de référence afin de créer un seul et unique filtre. Son expression est donc donnée par :

$$H_{COMP} = \sum_{i}^{n} a_i R_i, \tag{2.23}$$

où a_i est un coefficient permettant de pondérer l'importance de chaque image de référence dans la corrélation. L'avantage de cette technique est que les pics de corrélations des images de référence s'additionnent, rendant le filtre plus robuste aux effets de la rotation de l'image cible par exemple, permettant la reconnaissance d'un sujet sur une plus grande configuration du visage. En contrepartie, on observe une saturation du filtre lors de l'utilisation de nombreuses images de référence ou lorsque ces images sont très énergétiques [115].

2.4.2.2 Le filtre composite segmenté

Le filtre composite segmenté, dont le diagramme est présenté en figure 2.12, se présente comme une solution au problème de saturation du plan de corrélation du filtre composite basique. Ce résultat est obtenu en segmentant le plan de Fourier. Cela est effectué par un choix pixel par pixel suivant un critère de segmentation, qui peut être l'énergie du pixel, le gradient du spectre, le gradient de la phase ou bien la partie réelle du pixel.

Le choix du pixel se fait selon l'équation :

$$a_i Crit^i_{(u,v)} \leqslant a_j Crit^j_{(u,v)} \; \forall j \in [\![1,n]\!] \text{ et } j \neq i, \tag{2.24}$$

correspond au critère de segmentation du plan de Fourier. En utilisant par exemple comme critère de segmentation l'énergie de l'image, on compare l'énergie relative de chaque pixel après transformation de Fourier pour chaque image de référence avec l'ensemble des autres images de la base de référence. Le spectre

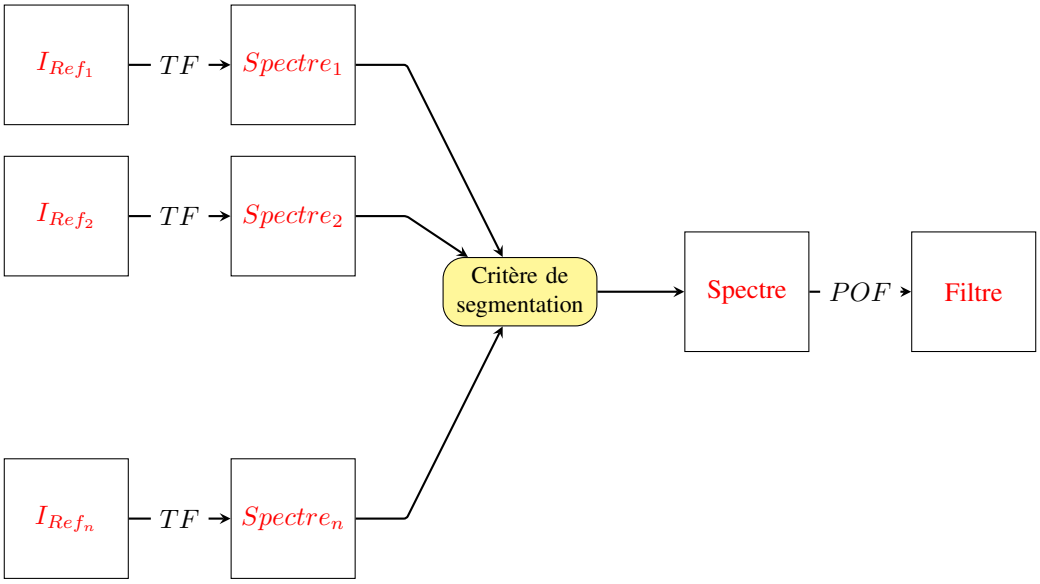


FIGURE 2.12 – Diagramme schématique du filtre composite segmenté.

résultant de cette méthode constitue donc celui regroupant les valeurs les plus énergétiques de la base pour chaque pixel. Le filtre, obtenu après sélection de la phase de l'image segmentée (Phase Only Filter), permet donc l'utilisation de la multi-corrélation sur un large set d'images de référence tout en évitant une saturation du plan de Fourier.

Différents critères de segmentation sont disponibles, à savoir l'énergie, le gradient complexe, le gradient de la phase, et la partie réelle de l'image.

Le critère énergétique est défini de la façon suivante : pour une position (u,v) donnée dans le plan de Fourier, on compare l'énergie du pixel pour l'ensemble des images de référence donnée par $E(u,v) = I^2(u,v)$. $I(u,v)$ est l'amplitude du pixel et $E(u,v)$ son énergie aux coordonnées (u,v). Afin d'éviter une segmentation faussée par une image trop énergétique, l'énergie de chaque pixel est normalisée par l'énergie totale de l'image dans le domaine spectral. On considère donc le pourcentage de l'énergie du pixel sur l'énergie de l'image complète. Le spectre du filtre composite segmenté est donc constitué des pixels les plus énergétiques de l'ensemble des images utilisées en référence. Le critère énergétique est donc donné par l'équation 2.25.

$$a_i \frac{E^i(u,v)}{\sum_{u=0}^{L}\sum_{v=0}^{H} E^i(u,v)} \geqslant \frac{E^j(u,v)}{\sum_{u=0}^{L}\sum_{v=0}^{H} E^j(u,v)}, \forall\, j \in [\![1,n]\!] \text{ et } j \neq i, \tag{2.25}$$

où $E^i(u,v)$ est l'énergie aux coordonnées (u,v) de la i$^{\text{ème}}$ image et $\sum_{u=0}^{L}\sum_{v=0}^{H} E^i(u,v)$, l'énergie de la i$^{\text{ème}}$ image. L'inconvénient majeur de ce critère est qu'il ne tient pas compte de la phase, bien qu'elle constitue la majeure partie de l'information contenue dans une image [105]. Malgré tout, il bénéficie de la forte dynamique engendrée par l'utilisation du module du spectre. Afin de prendre en compte l'information contenue dans la phase et ainsi remédier au problème posé par le critère énergétique de segmentation, on choisit d'utiliser le gradient complexe du spectre, donné par les dérivées partielles du spectre par rapport aux coordonnées u et v, selon l'équation 2.26.

$$\nabla S(u,v) = \begin{pmatrix} \frac{\partial S(u,v)}{\partial i} \\ \\ \frac{\partial S(u,v)}{\partial j} \end{pmatrix}, \tag{2.26}$$

où $S(u,v) = A(u,v)e^{i\varphi}$ est la valeur du spectre aux coordonnées (u,v) ($A(u,v)$ est le module du spectre et φ sa phase). Le module du gradient au carré est donné par :

$$|\nabla S(u,v)|^2 = \left|\frac{\partial S(u,v)}{\partial u}\right|^2 + \left|\frac{\partial S(u,v)}{\partial v}\right|^2, \tag{2.27}$$

mettant ainsi en évidence l'importance de la phase. La segmentation se fait donc maintenant selon le module du gradient complexe, suivant l'équation 2.28.

$$a_i|\nabla S^i(u,v)| \geqslant a_j|\nabla S^j(u,v)|, \forall\ j \in\ [\![1,n]\!] \text{ et } j \neq i. \tag{2.28}$$

Pour chaque pixel dans l'ensemble de la base de référence, celui retournant le plus fort gradient sera sélectionné pour la création du filtre composite segmenté. Contrairement au critère énergétique, la phase et le module du spectre sont tous deux pris en considération pour l'élaboration du filtre. Devant l'importance de l'information contenue dans la phase de l'image, il est également intéressant de s'intéresser à la variation de la phase, i.e. son gradient. Le gradient de la phase est donné par les dérivées partielles de la phase par rapport aux coordonnées de chaque pixel :

$$\nabla \varphi(u,v) = \begin{pmatrix} \frac{\partial \varphi(u,v)}{\partial i} \\ \\ \frac{\partial \varphi(u,v)}{\partial j} \end{pmatrix}, \tag{2.29}$$

où $\varphi(u,v)$ représente la phase aux coordonnées (u,v) . Le segmentation se faisant suivant le module du gradient, on obtient donc :

$$a_i|\nabla \varphi^i(u,v)| \geqslant a_j|\nabla \varphi^j(u,v)|, \forall\ j \in\ [\![1,n]\!] \text{ et } j \neq i. \tag{2.30}$$

À l'opposé du critère énergétique, le critère du gradient de la phase ne prend pas en compte le module du spectre et, par conséquent, bénéficie d'une faible dynamique. Le dernier critère de segmentation étudié est la partie réelle du pixel. L'avantage de cette méthode est qu'elle utilise à la fois l'information contenue dans l'amplitude et celle contenue dans la phase du spectre. La sélection du pixel se fait donc selon l'équation 27 :

$$a_i \frac{A^i(u,v)\cos^2(\varphi^i(u,v))}{\sum_{u=0}^{L}\sum_{v=0}^{H} E^i(u,v)} \geqslant \frac{A^j(u,v)\cos^2(\varphi^j(u,v))}{\sum_{u=0}^{L}\sum_{v=0}^{H} E^j(u,v)}, \forall\ j \in\ [\![1,n]\!] \text{ et } j \neq i, \tag{2.31}$$

De même que pour le critère énergétique, ce critère se fait après normalisation du pixel par l'énergie totale de l'image, afin d'éviter des problèmes dus aux différences d'énergie entre les images (engendrées par des différences d'exposition pendant la prise de vue). Contrairement au critère du gradient complexe, cette méthode donne une forte importance à l'information de la phase, tout en tenant compte du module.

2.4.3 Evaluation des performances

La corrélation de Vander Lugt dispose d'une kyrielle d'implantations possibles, allant d'un filtre POF simple, à une seule référence au filtre composite segmenté. Les différents filtres créés et testés seront explicités dans cette partie. Plusieurs critères de segmentation du filtre composite segmenté, à savoir l'énergie, le gradient, le gradient de la phase et la partie réelle, seront donc présentés, ainsi que les protocoles expérimentaux (i.e. les bases d'images cibles et de référence, les coefficients de pondération utilisés).

2.4.3.1 Le filtre de phase pure

Afin de déterminer les performances du Phase Only Filter, explicité en partie 2, la méthode a été testée sur les sujets 1 et 2 de la Pointing Head Pose Image Database (PHPID) [116]. 53 images recadrées à 215×215 pixels, variant de d'inclinaison 10 horizontalement ou verticalement entre chaque pose ont été conservées. La création du filtre a été effectuée en utilisant l'image de position centrale du visage du sujet 1. Ce filtre a ensuite été appliqué sur l'ensemble des images des sujets 1 et 2. De plus, afin de définir leur fidélité, la reconnaissance a été testée avec l'ensemble des critères de détection du pic de corrélation, à savoir le PCE, le PCE' , le PCE'' , le SNR et le SNR_{dB}.

La figure 2.13 présente la courbe PCE (Fig. 2.13a) et la courbe ROC (Fig. 2.13b) résultant de l'application du filtre POF sur les 2 personnes de la bases PHPID. La courbe ROC (Fig. 2.13b) présente 18% de vrais positifs pour un taux de fausse alarme à 0%. De même, on obtient un TPR à 58% pour 10% de FPR. Ces valeurs, relativement faibles, sont expliquées par l'observation de la courbe de distribution des PCE. En effet, la valeur du PCE pour l'image correspondant à celle utilisée en référence, correspondant à une forte corrélation (Fig. 2.13a), est très forte ($4{,}4.10^{-3}$) comparée aux valeurs du PCE des autres images du sujet 1 ($\sim\ 0{,}2.10^{-3}$). Ainsi, on obtient des valeurs de PCE pour la plupart des images cibles du sujet 1 très proches, voire même inférieures aux PCE des images du second sujet, introduisant des erreurs de classification. Le Phase Only Filter est donc très sensible aux rotations du visage, et par conséquent très discriminant au détriment de sa robustesse.

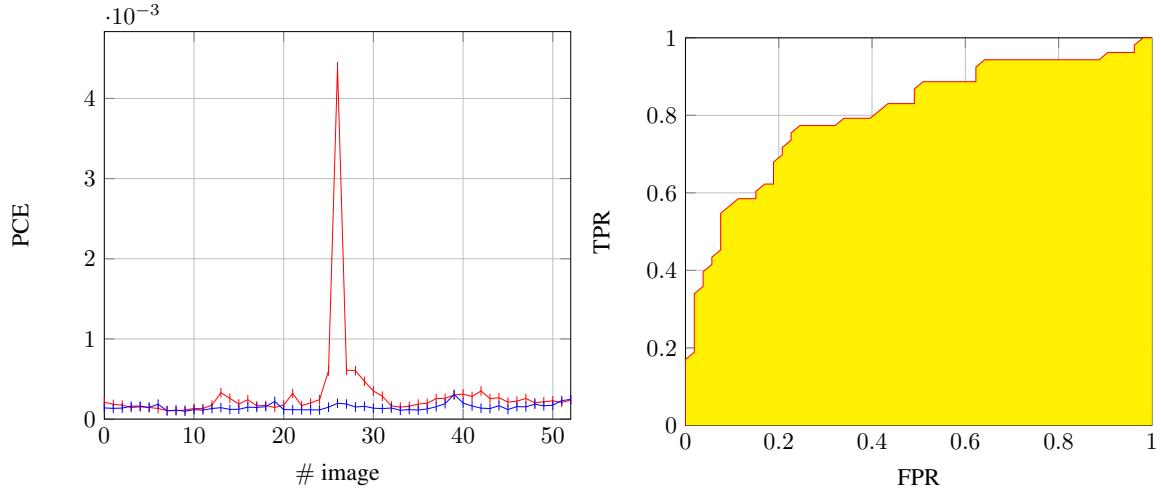


FIGURE 2.13 – Résultats obenus à l'aide du filtre POF : (a) courbes PCE, (b) courbe ROC.

Comme nous l'avons vu précédemment, différents critères de détection ont été proposés dans la littérature afin de détecter le pic de corrélation présent en sortie d'un corrélateur de Vander Lugt. Ces critères, à savoir le SNR, le SNR_dB , le PCE , le PCE' et le PCE'', ont donc été testés afin de déterminer leurs performances de détection du pic, et donc les améliorations apportées à la reconnaissance. Ainsi, un Phase Only Filter créé à partir de l'image centrale du sujet 1 de la base PHPID a donc été appliqué à la base d'images cibles. Les résultats, présentés en figures 2.14 et 2.15, ne montrent que très peu de différences. En effet, les courbes ROC (Fig. 2.14b, 2.14d, 2.14f, 2.15b et 2.15d) ont toutes la même tendance. Pour un taux de faux positifs de 0%, on obtient un taux de vrais positifs de 20% pour l'ensemble des critères. Un récapitulatif des résultats est donné en tableau 2.2. On observe également une très faible différence entre les critères, avec dans l'ensemble un

TPR à $5,.60\%$ pour un FPR à $7,55\%$. Une légère variation est également à noter pour le SNR, dont le TPR n'atteint $56,60\%$ que pour un FPR de $9,43\%$, et le PCE', avec un TPR et un FPR respectivement de $58,49\%$ et $7,55\%$. Le SNR est donc légèrement moins performant que les autres critères, le PCE' étant, lui, plus efficace, cependant sa résistance au bruit induit par le fond de l'image est plus faible que les autres critères [100].

TABLE 2.2 – Valeurs de TPR et FPR pour les différents critères de détection du pic de corrélation.

Critère	TPR(%)	FPR(%)
PCE	56,6	7,6
PCE'	58,5	7,6
PCE''	56,6	7.6
SNR	56,6	9,4
SNR_{dB}	56,6	7,6

Les différents critères de détection du pic de corrélation nous donnent des résultats sensiblement similaires. Bien que le PCE' nous retourne un taux de reconnaissance légèrement supérieur, celui-ci est intrinsèquement plus sensiblement aux variations du fond de l'image. Du fait de cette faible différence des résultats obtenus suivant le critère de détection du pic de corrélation, nous choisissons par le suite le critère le plus largement utilisé dans la litérature, le PCE.

2.4.3.2 Fitre composite segmenté

Le filtre composite segmenté, créé à partir des images 9, 27 et 43 du sujet de référence a été appliqué à l'ensemble des images cibles de la base, le critère de segmentation étant la partie réelle. Comme cela a été observé précédemment, les valeurs du PCE pour l'ensemble des images cibles est souvent très faibles par rapport à la référence centrale (image 27). Plusieurs valeurs du PCE correspondant aux images cibles de la personne de référence sont par conséquent inférieures aux valeurs du PCE des images cibles du second sujet, augmentant le taux de fausses alarmes et réduisant de par le fait les performances du filtre. L'idée est donc d'augmenter la robustesse du filtre en introduisant deux autres images de référence. La figure 2.16 présente les courbes PCE (Fig. 2.16a) et ROC (Fig. 2.16b) obtenues à l'aide d'un filtre composite segmenté utilisant 3 images de référence. On observe une amélioration de la reconnaissance par rapport au filtre de phase pure. En effet, le filtre atteint maintenant un taux de reconnaissance de 34% pour 0% de fausse alarme, alors qu'il était de 20% pour le même taux de fausse alarme pour le filtre POF. Ceci s'explique par une augmentation du PCE correspondant aux images de référence additionnelles et aux images proches de ces dernières (avec $+/-10^{\circ}$ de rotation du visage), comme on peut l'observer les courbes de distribution du PCE (Fig. 2.16a). L'introduction d'images de références permet donc une amélioration certaine des performances du classifieur. De plus, le temps de corrélation reste inchangé, une seule corrélation étant toujours nécessaire, bien que le temps de création du filtre soit légèrement plus élevé.

2.4.3.2.1 Effet du choix des références Comme on vient de l'observer, ajouter des références au corrélateur de Vander Lugt permet d'améliorer les performances du classifieur. Afin de reconnaitre la majeure partie de la base, et ainsi d'obtenir une classification quasi-parfaite, l'idée est donc d'augmenter le nombre d'images de référence utilisées dans la création du filtre composite segmenté. Pour ce faire, on a remarqué que 13 images de référence suffisaient à permettre une reconnaissance d'une grande partie de la base, choisies afin d'optimiser la reconnaissance, par méthode de test et erreur. Différentes implémentations de cette corrélation à 13 références ont donc été évaluées. Il est en effet possible de créer soit 1 seul filtre segmenté à 13 références, soit plusieurs filtres contenant chacun une partie de ces 13 références. Les filtres ont été créés

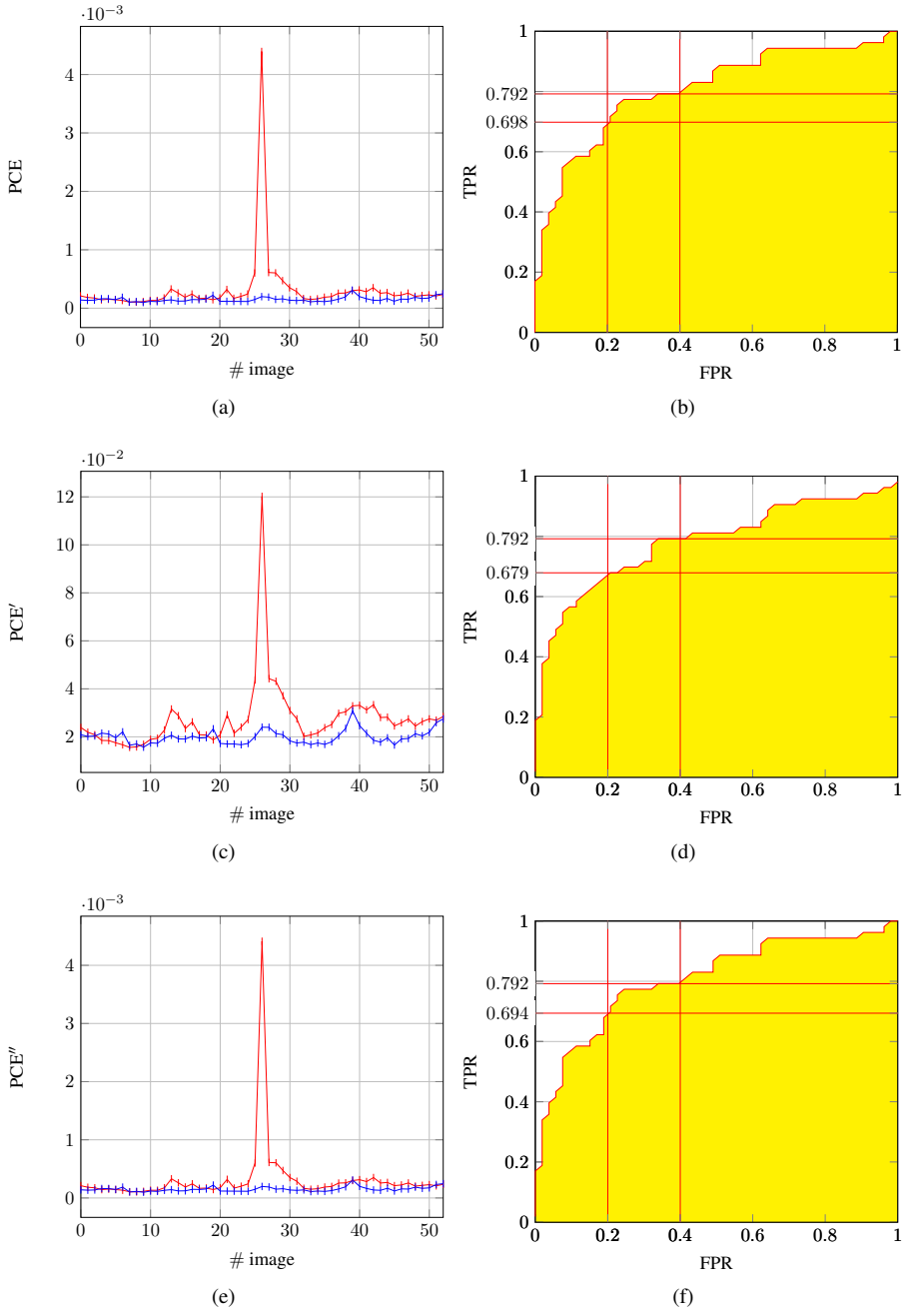


FIGURE 2.14 – Résultats obtenus à l'aide du filtre POF et les métriques de détection du pic PCE, PCE' et CE'' : (a) PCE, (c) PCE', (e) PCE'' ; (b), (d), (f), les courbes ROC correspondantes..

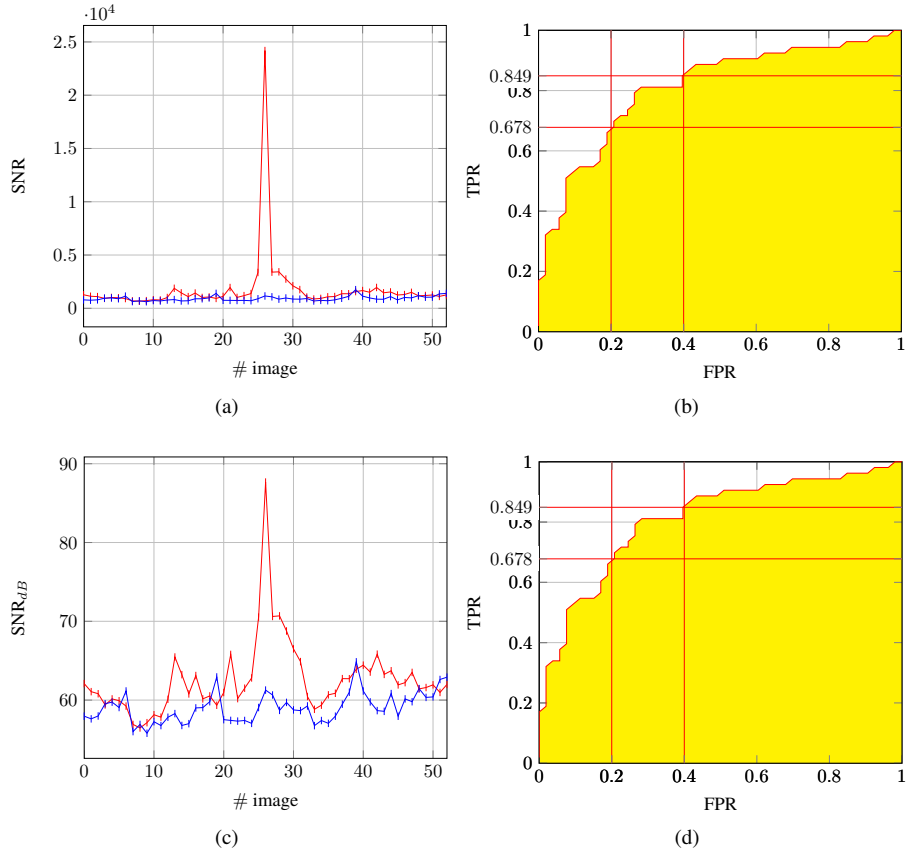
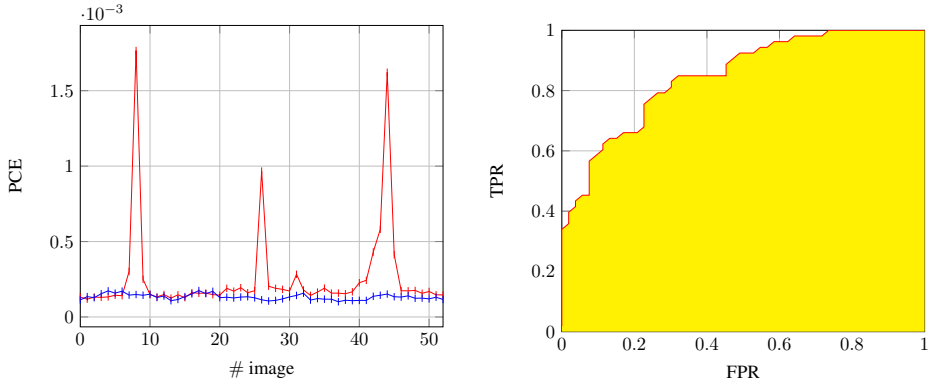


FIGURE 2.15 – Résultats obtenus à l'aide du filtre POF et les métriques de détection du pic SNR et SNR_{dB} : (a) SNR, (c) $SNRdB$; (b) et (d) les courbes ROC correspondantes.

utilisant les images de la personne 1 de la base PHPID et ont été segmentés à l'aide du critère énergétique et un coefficient de pondération de 4 sur l'image centrale. Différentes implémentations ont donc été testées, à savoir l'utilisation de 6 filtres à 3 références (Fig. 2.17a et 2.17b), 3 filtres à 5 références (Fig. 2.17c et 2.17d), 2 filtres à 7 références (Fig. 2.17e et 2.17f) et enfin 1 seul filtre à 13 références (Fig. 2.17g et 2.17h) (l'image 27 de position centrale étant systématiquement utilisée dans la construction de chacun des filtres, afin de privilégier la reconnaissance de face). Les résultats, résumés sur le tableau 2.3, présentent une nette amélioration des performances par rapport à l'expérience précédente, utilisant uniquement 3 images de référence. En effet, on observe pour 0% de fausse alarme un TPR de 62% au minimum (utilisation de 3 filtres). L'utilisation de 2 filtres à 7 références engendre les meilleurs résultats, avec un TPR de 83% et 89% pour 0% et 10% de fausse alarme, respectivement. Les performances plus faibles du filtre à 13 références sont expliquées par un effet de saturation, dû à l'utilisation d'un trop grand nombre de références. L'avantage de l'utilisation de la combinaison de deux filtres à sept références, en plus de son efficacité légèrement supérieure, est qu'elle nécessite deux corrélations, tandis que la combinaison de 6 filtres en nécessite six. Elle constitue donc le meilleur compromis entre une reconnaissance optimale et un faible temps computationnel.

2.4.3.2.2 Effet du critère de segmentation Différents critères de segmentation sont disponibles pour la construction du filtre composite segmenté, utilisant plus ou moins l'information contenue dans la phase ou la



(a) (b)

FIGURE 2.16 – Résultats obtenus à l'aide d'un filtre segmenté à 3 références : (a) courbes PCE ; (b) courbe ROC.

TABLE 2.3 – Valeurs de TPR et FPR pour les différentes combinaison de filtres composites segmentés pour 13 références.

Nombre de filtres	TPR(%)	FPR(%)
6	79%	0%
	89%	10%
3	62%	0%
	79%	10%
2	83%	0%
	89%	10%
1	77%	0%
	83%	10%

dynamique du module du spectre des images de référence. Les critères évalués en figure 2.18, sont l'énergie du spectre (Fig. 2.18a et 2.18b), son gradient complexe (Fig. 2.18c et 2.18d), son gradient de phase (Fig. 2.18e et 2.18f) ou encore sa partie réelle (Fig. 2.18g et 2.18h). Les résultats sont résumés en tableau 2.4.

On observe que le critère de segmentation influe fortement sur le comportement du filtre. Le gradient de la phase, bien qu'ayant un TPR faible à 0% de FPR effectue une forte augmentation de son taux de reconnaissance à 10%, contrairement aux critères tels que l'énergie et la partie réelle. Ce comportement est expliqué par la grande importance donnée à la phase, contenant la majeure partie de l'information. Le critère le plus efficace, donnant les meilleurs résultats, est clairement le gradient complexe. En effet, ce critère bénéficie à la fois des informations du module et de la phase, lui fournissant ainsi une forte robustesse tout en maintenant la grande discrimination du filtre POF. L'utilisation du gradient complexe permet des résultats de l'ordre de ceux obtenus avec le critère énergétique lors de l'utilisation de 13 images de référence, à savoir un TPR respectif de 60% et 70% pour 0% et 10% de fausse alarme.

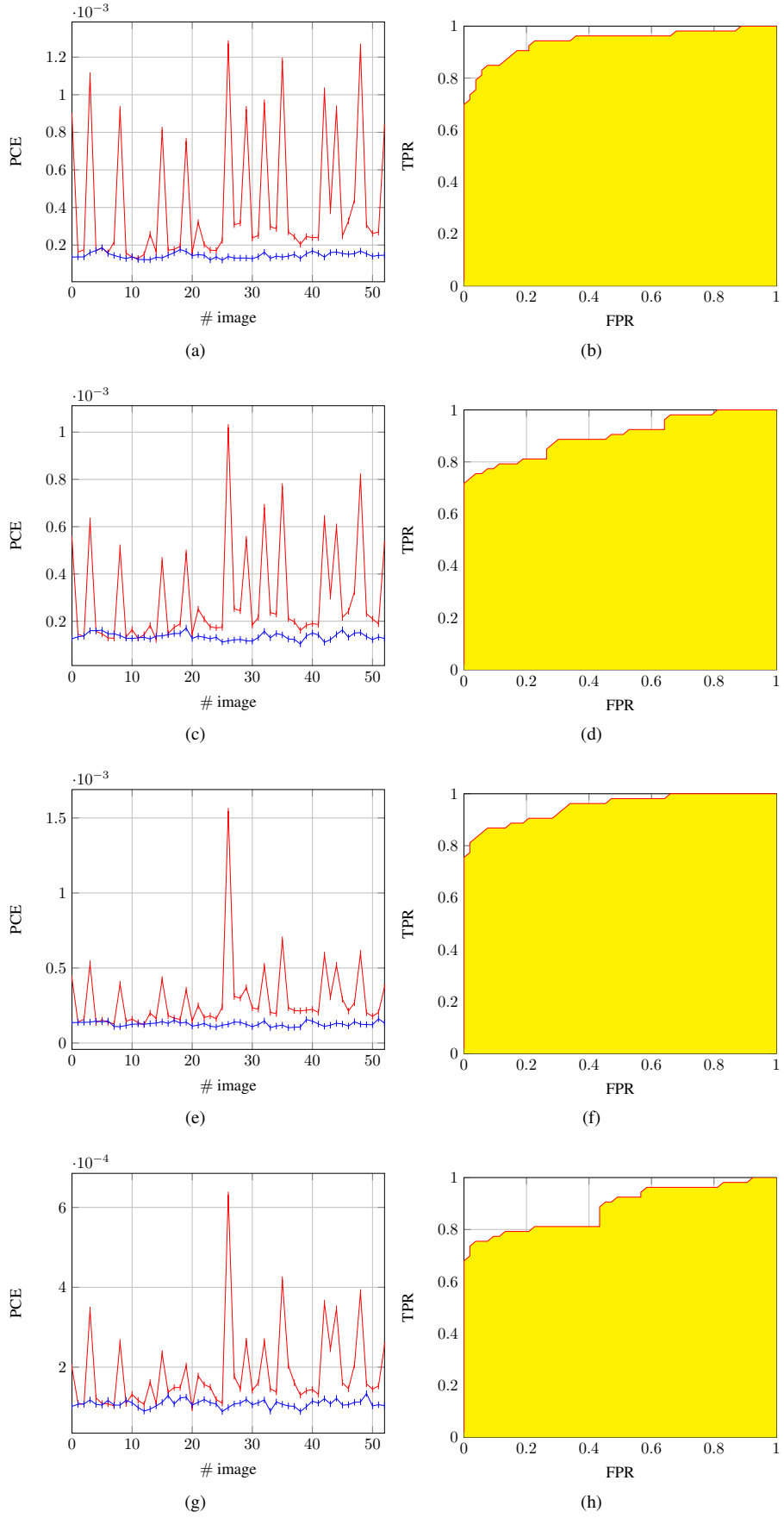


FIGURE 2.17 – Courbes PCE et ROC obtenues à l'aide d'un filtre segmenté : (a) et (b) à l'aide de 6 filtres à 3 références ; (c) et (d) à l'aide de 3 filtres à 5 références ; (e) et (f) à l'aide de 2 filtres à 7 références ; (g) et (h) à l'aide de 1 filtre à 13 références.

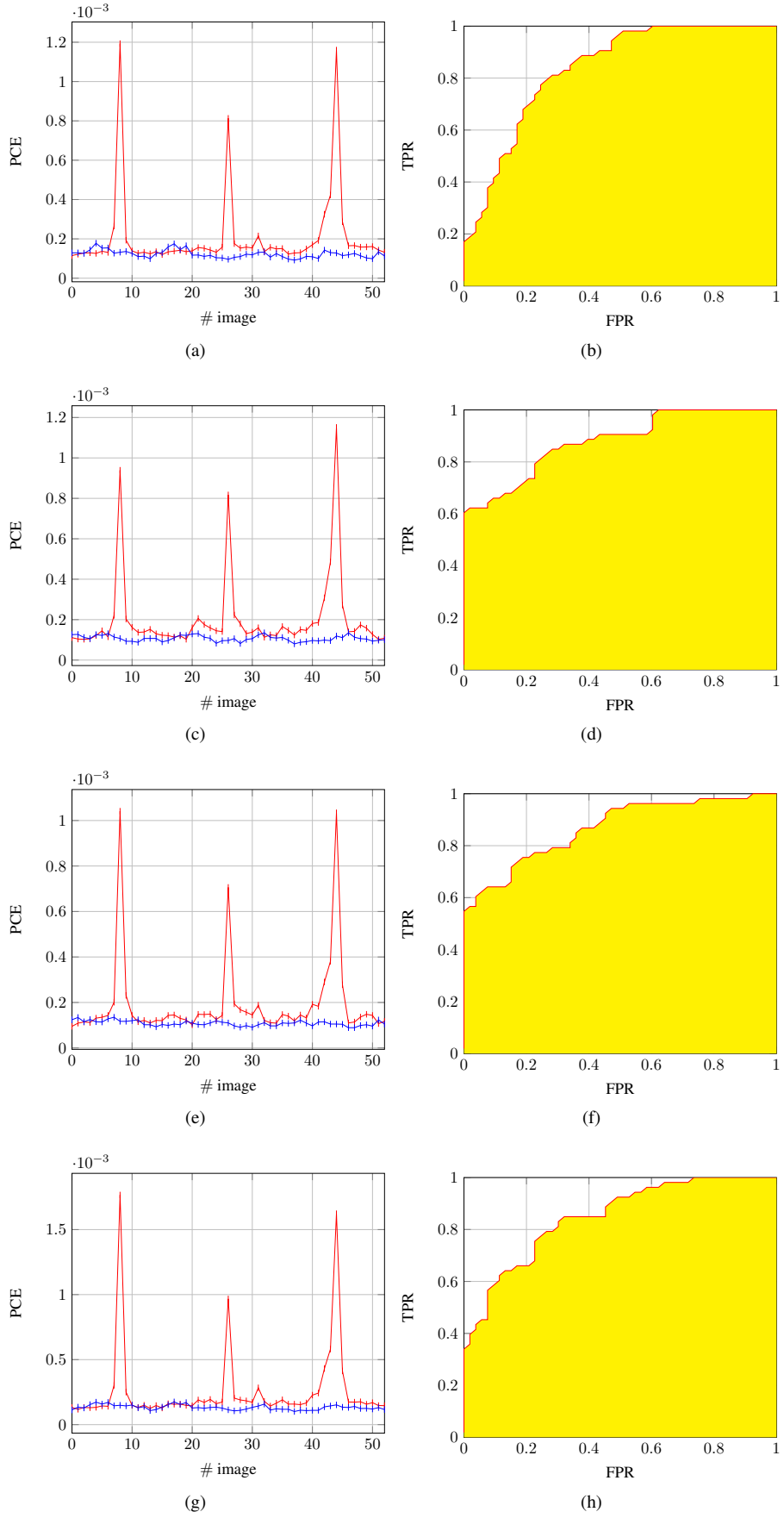


FIGURE 2.18 – Courbes PCE et ROC obtenues à l'aide d'un filtre segmenté à 3 références : (a) et (b) critère énergétique ; (c) et (d) critère du gradient complexe ; (e) et (f) critère du gradient de la phase ; (g) et (h) critère de la partie réelle.

TABLE 2.4 – Valeurs de TPR et FPR pour les différents critères de segmentation.

Critère	TPR(%)	FPR(%)
Energie	22%	0%
	40%	10%
Gradient complexe	60%	0%
	70%	10%
Gradient de la phase	32%	0%
	59%	10%
Partie réelle	21%	0%
	39%	10%

2.5 Le Joint Transform Correlator

Aboutissant le travail de Weaver et Goodman sur la convolution optique de deux images, la méthode de corrélation à transformée de Fourier conjointe (JTC) a été introduite en 1966 [117] afin de résoudre certaines limitations importantes du corrélateur de Vander Lugt. Tout comme le corrélateur de Vander Lugt, le JTC est basé sur l'architecture optique dite "4f".

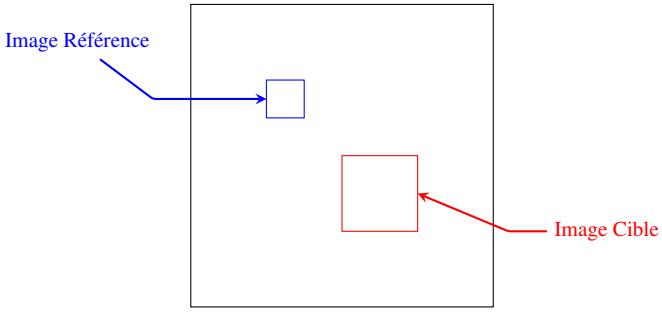


FIGURE 2.19 – Création du plan d'entrée.

Contrairement au corrélateur de Vander Lugt, le JTC dispose d'un plan d'entrée comprenant à la fois l'image référence et l'image cible. Un exemple de positionnement des images dans le plan d'entrée est présenté en Figure 2.19. L'étape critique de la création du plan d'entrée repose sur le positionnement des images cible et référence. En effet, deux images distantes d'une distance h sur le plan d'entrée engendreront des pics d'inter-corrélation distants de $H = 2 \times h$. Le plan d'entrée doit donc être constitué de façon à respecter cette caractéristique.

En posant (x,y) les coordonnées de l'image référence i_{Ref} et l'image référence et i_{Cible} l'image cible, nous obtenons le plan d'entrée P (Eq. 2.32) :

$$P(x,y) = i_{Ref}(x,y) + i_{Cible}(x + h_x, y + h_y). \tag{2.32}$$

Une transformée de Fourier du plan d'entrée nous retourne finalement le spectre joint, ou plan de Fourier, donné par l'équation 2.33.

$$t(u,v) = \begin{array}{l} |I_{Ref}(u,v)| \times \exp(\varphi_{I_{Ref}}(u,v) + \\ |I_{Cible}(u,v)| \times \exp(\varphi_{I_{Cible}}(u,v) \times \exp(-j(uh_x + vh_y)) \end{array} \tag{2.33}$$

Finalement, une transformée de Fourier inverse du module au carré du spectre joint $|t(u,v)|^2$ nous permet d'obtenir le plan de corrélation, contenant trois principaux pics : le pic central d'autocorrélation, c'est-à-dire à la somme de la corrélation de l'image référence avec elle-même et de l'image cible avec elle-même ; et les pics d'intercorrélation, symétriquement disposés par rapport au pic d'autocorrélation, correspondant à la corrélation des images cible et référence entre elles. Le synoptique du JTC est présenté en figure 2.20. Comme nous le verrons au chapitre 4, la position des pics d'intercorrélation est conditionnée par la position des images cible et référence sur le plan d'entrée.

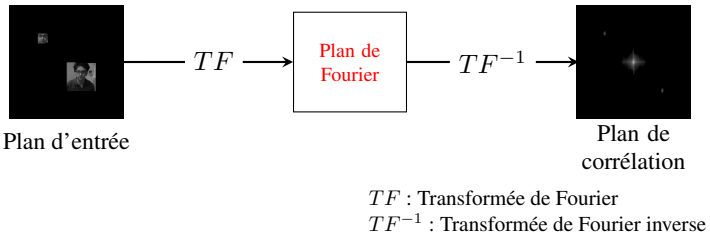


FIGURE 2.20 – Diagramme schématique du corrélateur à spectre joint.

Différentes versions et améliorations du JTC sont présentes dans la littérature, nous en présentons ici quelques-unes.

2.5.1 Le JTC classique

Le JTC dit "classique" est l'application directe de ce qui vient d'être énoncé précédemment. L'intensité du spectre joint est utilisée pour la création du plan de corrélation. Celle-ci est exprimée par l'équation 2.34 :

$$E(u,v) = |t(u,v)|^2 = \begin{cases} |I_{Ref}(u,v)|^2 + |I_{Cible}(u,v)|^2 \\ +|I_{Ref}(u,v)|e^{\phi_s(u,v)}|I_{Cible}(u,v)|e^{-\phi_r(u,v)+j(ul+vl)} \\ +|I_{Ref}(u,v)|e^{-\phi_s(u,v)}|I_{Cible}(u,v)|e^{\phi_r(u,v)-j(ul+vl)} \end{cases} \tag{2.34}$$

où $t_(u,v)$ est l'intensité du spectre joint. Une transformée de Fourier inverse de $E(u,v)$ nous retourne le plan de corrélation. L'équation 2.34 se décompose en : (i) $|I_{Ref}(u,v)|^2 + |I_{Cible}(u,v)|^2$, la somme des autocorrélations des images cible et référence ; (ii) $|I_{Ref}(u,v)|e^{\phi_s(u,v)}|I_{Cible}(u,v)|e^{-\phi_r(u,v)+j(ul+vl)}$ la corrélation de l'image référence avec l'image cible ; (iii) $|I_{Ref}(u,v)|e^{-\phi_s(u,v)}|I_{Cible}(u,v)|e^{\phi_r(u,v)-j(ul+vl)}$ la corrélation de l'image cible avec l'image référence. L'intensité des pics d'intercorrélation est conditionnée par le degré de similarité entre les images référence et cible.

Un exemple de plan de corrélation est présenté en figure 2.21. Nous pouvons observer deux pics d'intercorrélation larges et évasés et un pic d'autocorrélation beaucoup plus intense. Ces deux caractéristiques représentent la limitation majeure du JTC classique : en effet, la largeur et l'intensité du pic d'autocorrélation, appelé "ordre zéro", peut engendrer une impossibilité de détection des pics d'intercorrélation en cas de faible ressemblance avec les images. En effet, en cas de faible ressemblance, les pics d'intercorrélation ont une intensité très faible et leur évasement peut donc entrainer une disparition de ces pics dans l'ordre zéro.

2.5.2 Le JTC sans ordre zéro

Afin de pallier au problème causé par la présence d'un pic d'autocorrélation intense et large, rendant le système sensible au bruit, notamment en cas de faible distance entre les images cible et référence sur le plan d'entrée, Li et al. ont proposé en 1998 [118] une suppression mathématique de l'ordre zéro du plan de corrélation. Comme nous venons de le voir, la première ligne de l'équation 2.34 correspond à la somme

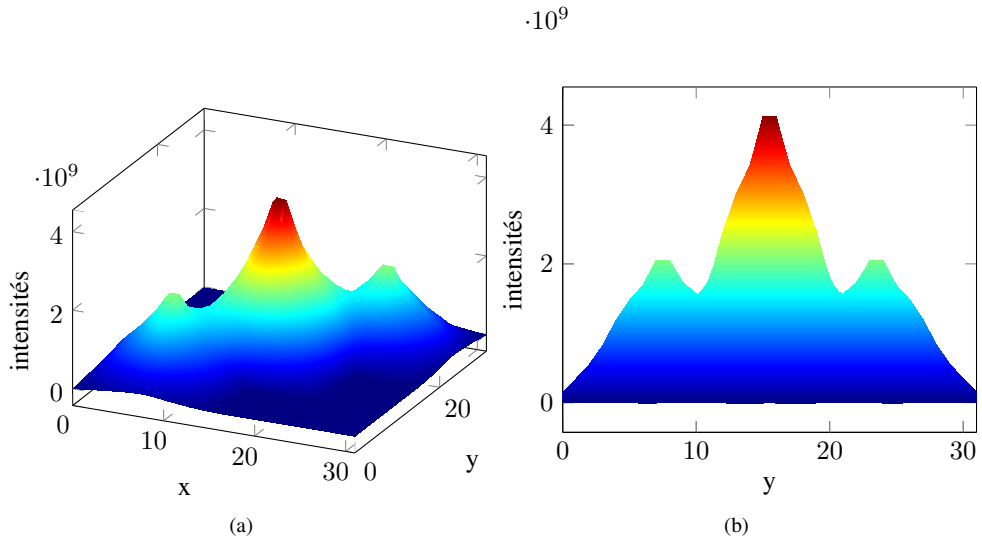


FIGURE 2.21 – Plan de corrélation du JTC Classique : (a) vue globale ; (b) vue suivant y.

des autocorrélations des images cible et référence, c'est-à-dire à l'ordre zéro. Le principe de leur approche, présenté schématiquement en figure 2.22, consiste à calculer séparément les intensités d'autocorrélation pour les retrancher du plan de corrélation.

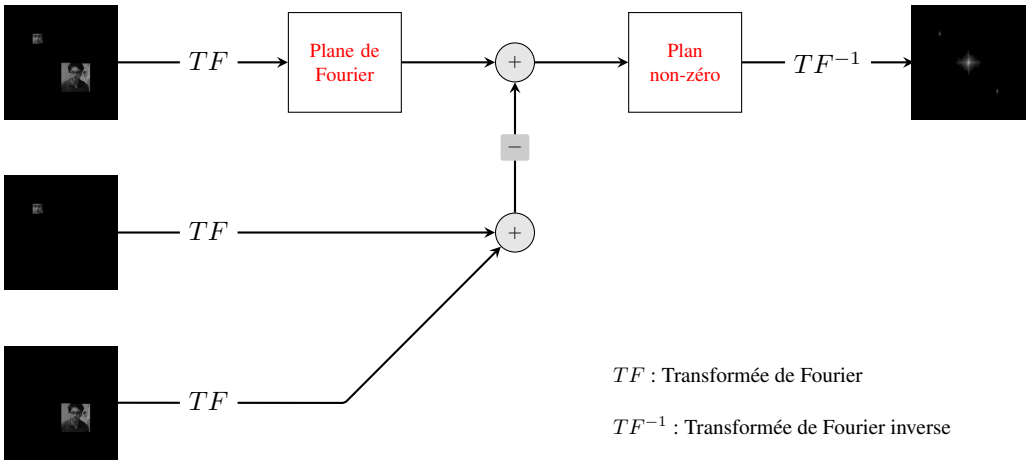


FIGURE 2.22 – Corrélateur à spectre joint sans ordre zéro.

Pour ce faire, deux plans d'entrée supplémentaires sont créés, l'un contenant uniquement l'image cible, l'autre l'image référence. Leur transformée de Fourier est calculée indépendamment et leur intensité est retranchée du spectre joint (Eq. 2.35) :

$$E_{nz}(u,v) = E(u,v) - |I_{Ref}(u,v)|^2 - |I_{Cible}(u,v)|^2, \tag{2.35}$$

où $E_{nz}(u,v)$ est le spectre joint sans ordre zéro. Ainsi, l'équation 2.34 devient (Eq. 2.36) :

$$E_{nz}(u,v) = \begin{cases} +|I_{Ref}(u,v)|e^{\phi_s(u,v)}|I_{Cible}(u,v)|e^{-\phi_r(u,v)+j(uh+vh)} \\ +|I_{Ref}(u,v)|e^{-\phi_s(u,v)}|I_{Cible}(u,v)|e^{\phi_r(u,v)-j(uh+vh)} \end{cases} \quad (2.36)$$

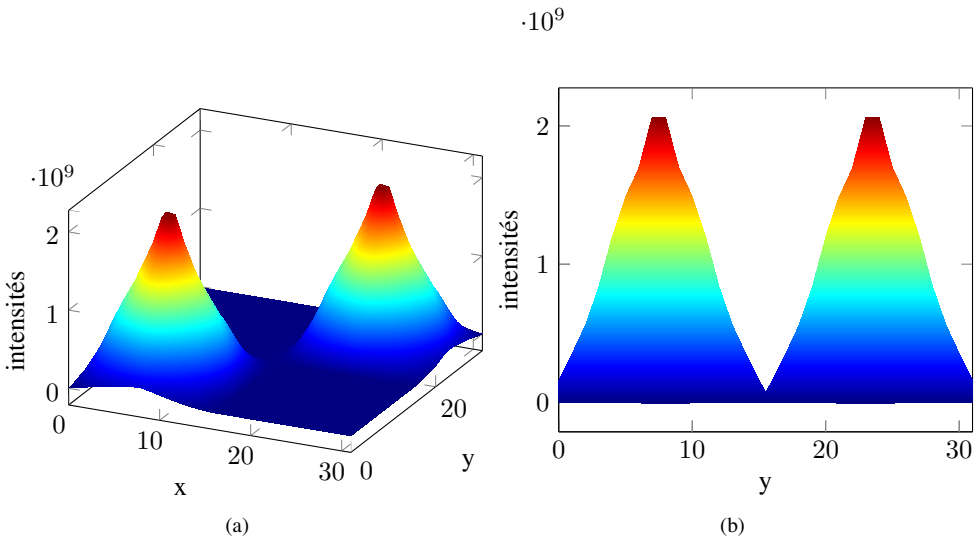


FIGURE 2.23 – Plan de corrélation du JTC sans ordre zéro : (a) vue globale ; (b) vue suivant y.

L'application de l'algorithme de suppression de l'ordre zéro est présentée en figure 2.23. Comme nous pouvons le remarquer, le pic central d'autocorrélation a été correctement éliminé. Les risques de chevauchement des pics d'inter et d'autocorrélations sont donc dorénavant écartés, malgré tout, le problème de faible intensité et d'évasement des pics de corrélation demeurent.

2.5.3 Le JTC non-linéaire

Finalement, afin de contrôler la sensibilité du corrélateur, et donc de résoudre le problème induit par les pics de corrélation larges et de faible intensité, [119] proposa en 1989 l'introdution d'une fonction de linéarité dans le spectre joint. Cela est effectué en l'élevant à une puissance k suivant l'équation 2.37 :

$$E_{nl}(u,v) = g[E(u,v)] = E^k(u,v) = (|t(u,v)|^2)^k,\ k \in [0,1], \quad (2.37)$$

où g est la fonction de non-linéarité. Le coefficient k permet de régler le meilleur compromis pour une application donnée entre la robustesse et la discrimination.

La figure 2.24 présente les plans de corrélation du JTC non-linéaire en vue globale et suivant l'axe y pour différentes valeurs de coefficient de non-linéarité k. Le pic d'autocorrélation a été supprimé à l'aide de la méthode décrite précédemment. Les figures 2.24a et 2.24b ont été obtenues pour un coefficient de non-linéarité $k = 0{,}3$, 2.24c et 2.24d pour $k = 0{,}5$ et 2.24e et 2.24f pour $k = 0{,}7$. Comme nous pouvons le constater, le coefficient de non-linéarité influence profondément le comportement du corrélateur. En effet, on observe des pics de corrélation larges et peu intenses par comparaison au bruit présent dans le plan de corrélation pour un coefficient élevé $k = 0{,}7$ et des pics extrêmement fins pour un coefficient faible $k = 0{,}3$. L'utilisation d'une valeur $k = 0.5$ présente un compromis entre les deux valeurs présentées précédemment. Un coefficient élevé, proche de 1, engendre des pics larges et peu intenses par comparaison au reste du plan de corrélation,

tandis qu'un coefficient faible, proche de 0 présentera des pics très localisés et intenses par rapport au reste du plan. Finalement, l'utilisation de $k = 1$ entrainera un corrélateur JTC classique, tandis qu'on obtiendra un JTC binaire avec $k = 0$. Par conséquent, ceci aura un effet important sur le comportement du corrélateur : de larges pics de corrélation permettent d'obtenir un corrélateur robuste quant aux variations présentes entre la cible et la référence tandis que des pics localisés engendrent un corrélateur très discriminant. Par ailleurs, comme cela a été signifié, il existe une relation entre la position des pics de corrélation sur le plan et la position relative des images cible et référence sur le plan d'entrée. Des pics peu larges permettront donc une localisation beaucoup plus précise de l'image cible lors d'une utilisation du JTC dans un système de suivi. Il s'agit donc detrouver un compromis entre la robustesse et la discimination du corrélateur, tout en prenant en compte, si besoin est, les performances de localisation.

2.5.4 Effets des paramètres sur le comportement du corrélateur à spectre joint

L'introduction d'une fonction de non-linéarité et la suppression de l'ordre zéro ont la capacité d'influer significativement sur le comportement du corrélateur. L'effet du coefficient de non-linéarité sur le JTC est illustré par les tableaux 2.5 pour un JTC classique et 2.6 pour un JTC sans ordre zéro. Nous présentons le plan de corrélation et le PCE pour une auto-corrélation (images cible et référence identiques), une corrélation vraie (images cibles et référence différentes mais issues du même objet) et une corrélation fausse (images cible et référence issues de deux objets différents).

Nous observons dans un premier temps l'effet de la suppression de l'ordre zéro pour un JTC sans fonction de non-linéarité ($k = 1$). Le pic d'auto-corrélation a en effet bien été retiré du plan de corrélation. Cependant, pour le JTC classique et le JTC sans ordre zéro, nous obtenons une valeur de PCE plus élevée pour une corrélation fausse que pour une corrélation vraie ou une auto corrélation. Toutefois, les valeurs de PCE sont globalement plus élevées lorsque le pic d'auto-corrélation a été supprimé.

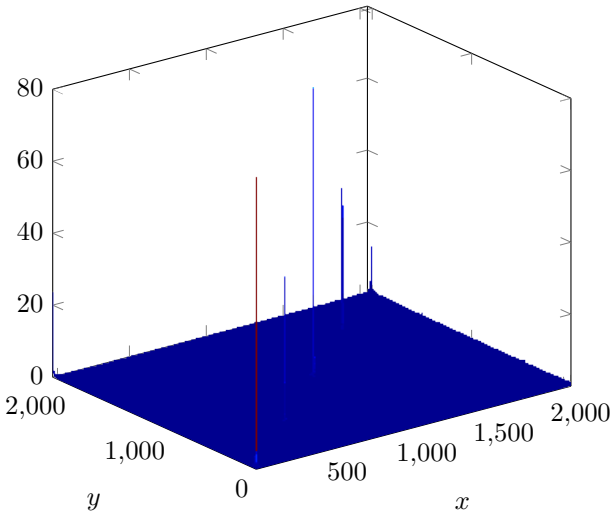
Comme nous l'avons précédemment exposé, le coefficient de non-linéarité permet d'influer sur la discrimination et la robustesse du corrélateur. Nous retrouvons bien ce comportement ici. Le choix d'un coefficient $k = 0{,}8$, proche du JTC classique, permet une discrimination suffisante pour différencier la corrélation vraie de la corrélation fausse, tout en étant suffisamment robuste pour retrouver des valeurs similaires pour la corrélation vraie et l'auto-corrélation. À l'inverse, un coefficient $k = 0{,}2$ engendre un PCE similaire pour la corrélation fausse et la corrélation vraie, (de l'ordre de 1.10^{-2}), et un PCE élevé (de l'ordre de 1.10^{-1}) pour l'auto-corrélation. Ce coefficient nous donne donc bien un corrélateur extrêmement discriminant, corrélant uniquement lorsque les images référence et cible sont identiques. Nous retrouvons ces résultats pour le corrélateur sans ordre zéro.

La fonction de non-linéarité a donc un rôle extrêmement important dans l'adaptation de l'architecture à l'application voulue. En effet, celui-ci permet de modifier radicalement le compromis entre la discrimination et la robustesse. La suppression du pic d'auto corrélation, quant-à-elle, a un rôle secondaire, permettant uniquement un recouvrement des pics d'inter-corrélation par le pic d'auto corrélation.

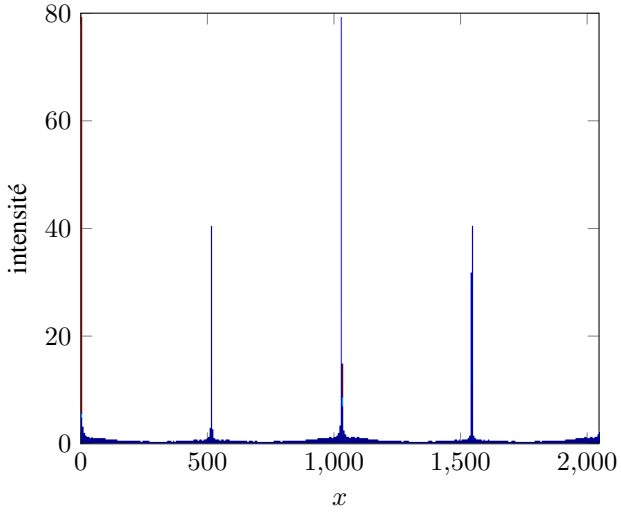
2.6 Discussion

Les deux architectures majeures pour la corrélation, bien qu'étant basées toutes deux sur le montage optique "4f", diffèrent sensiblement quant à leurs performances et à leurs applications potentielles.

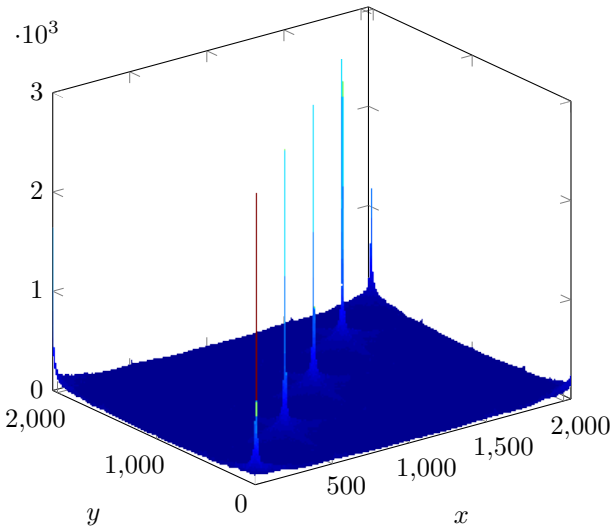
Le corrélateur de Vander Lugt se distingue par une création d'un filtre, indépendamment de l'étape de corrélation. Cette propriété permet la réalisation d'un filtre optimisé pour l'application voulue. Ainsi, ce type de corrélateur est adapté à la reconnaissance d'objet, car il est à même d'utiliser plusieurs filtres préétablis pour le même objet et donc d'augmenter les capacités de reconnaissance. En effet, alors que le filtre adapté permet d'obtenir un filtre très robuste, le filtre de phase pure, quant-à-lui, induit une corrélation très discriminante, optimisant au maximum l'auto-corrélation. Ces deux filtres sont problématiques lorsque l'on souhaite recon-



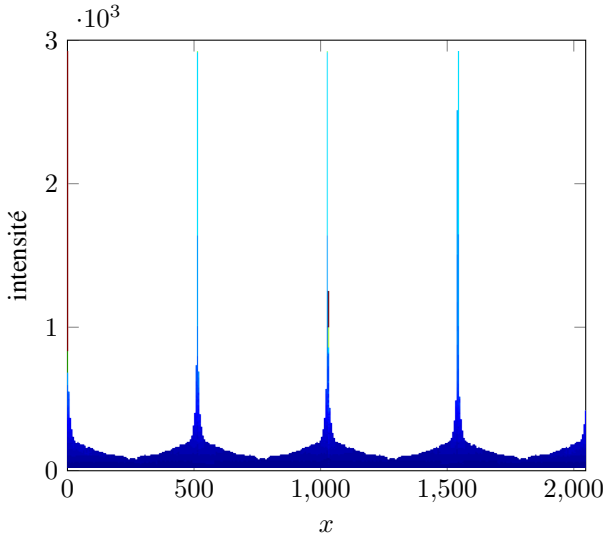
(a)



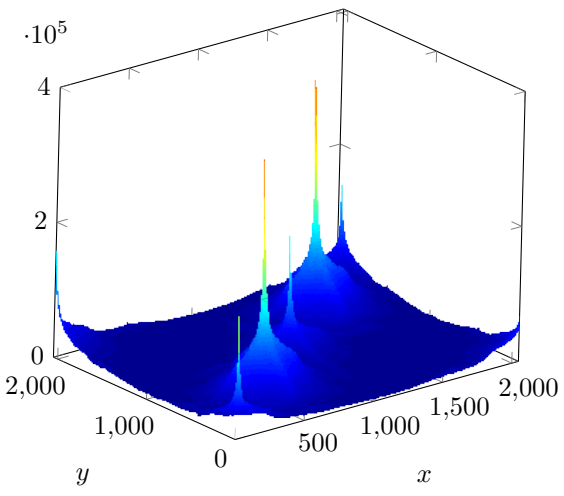
(b)



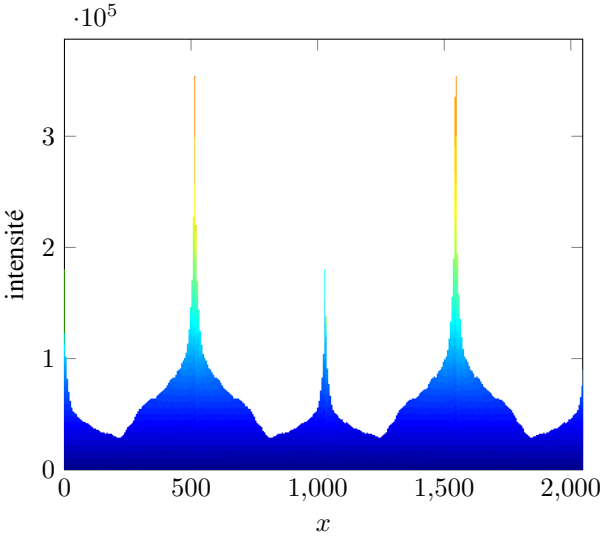
(c)



(d)



(e)



(f)

FIGURE 2.24 – Plan de corrélation du JTC non-linéaire sans ordre zéro avec différentes valeurs de coefficient de non-linéarité k : (a) et (b) $k = 0{,}3$; (c) et (d) $k = 0{,}5$; (e) et (f) $k = 0{,}7$; (a), (c)et (e) vue globale ; (b), (d) et (f) vue suivant y.

TABLE 2.5 – Plans de corrélation et valeurs de PCE du JTC non-linéaire suivant différentes valeurs de coefficient de non-linéarité, pour une autocorrélation, une corrélation vraie (images différentes du même objet) et une corrélation fausse.

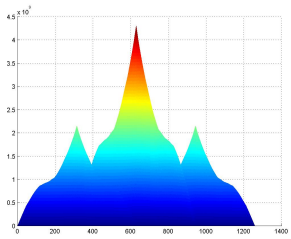
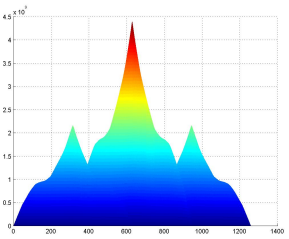
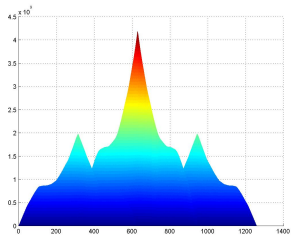
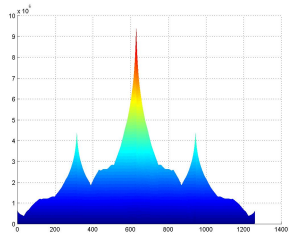
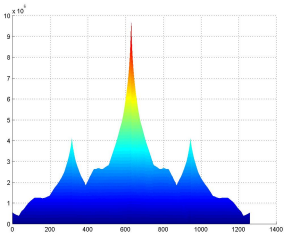
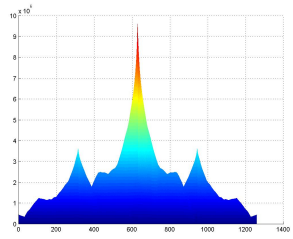
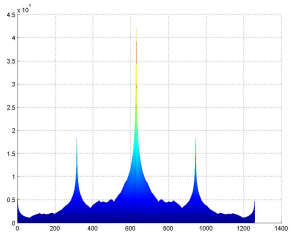
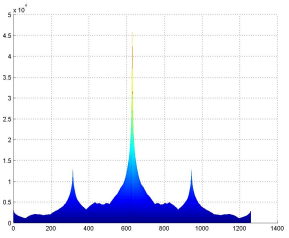
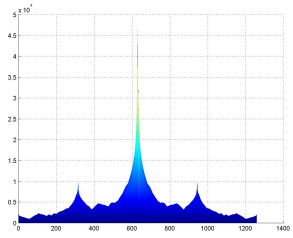
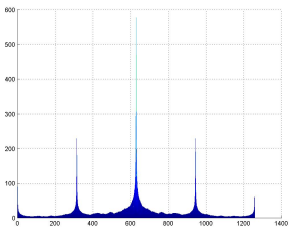
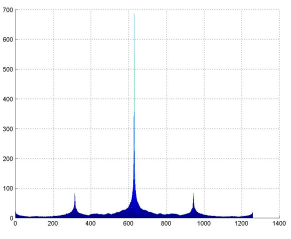
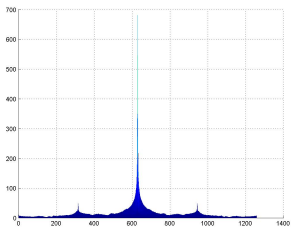
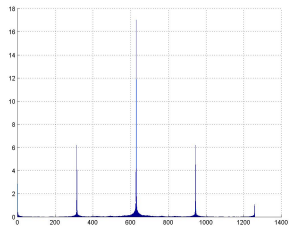
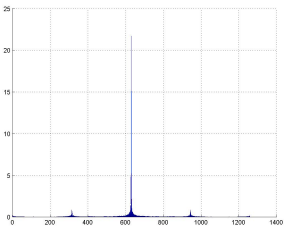
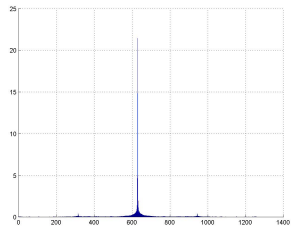
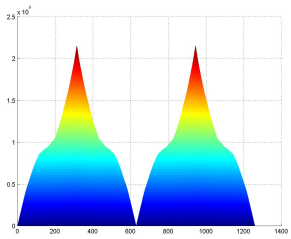
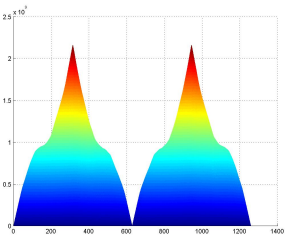
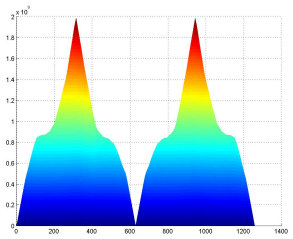
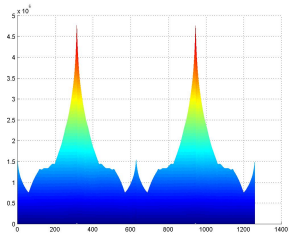
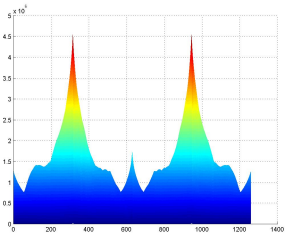
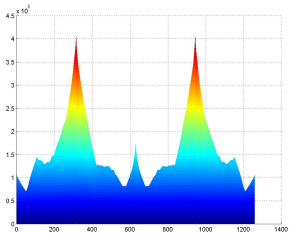
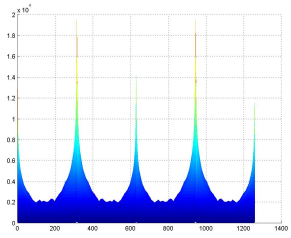
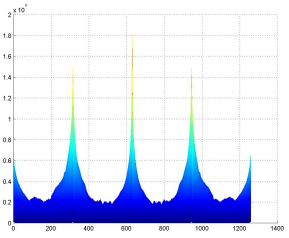
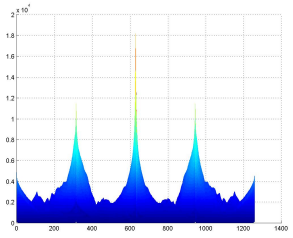
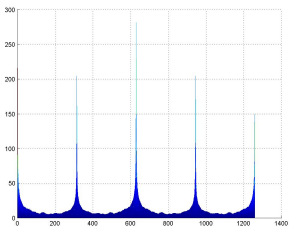
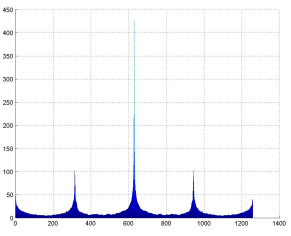
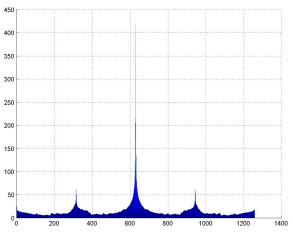
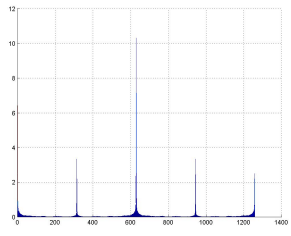
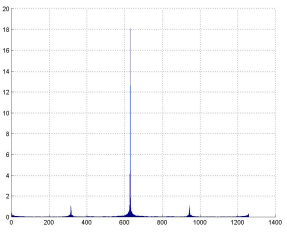
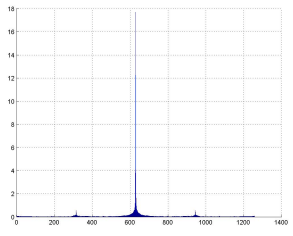
k	Autocorrélation	Corrélation vraie	Corrélation fausse
1	 $PCE = 4{,}521.10^{-5}$	 $PCE = 4{,}444.10^{-5}$	 $PCE = 4{,}539.10^{-5}$
0,8	 $PCE = 1{,}160.10^{-4}$	 $PCE = 1{,}004.10^{-4}$	 $PCE = 9{,}272.10^{-5}$
0,6	 $PCE = 1{,}100.10^{-3}$	 $PCE = 5{,}403.10^{-4}$	 $PCE = 3{,}615.10^{-4}$
0,4	 $PCE = 5{,}355.10^{-2}$	 $PCE = 8{,}989.10^{-3}$	 $PCE = 3{,}380.10^{-3}$
0,2	 $PCE = 2{,}911.10^{-1}$	 $PCE = 7{,}289.10^{-2}$	 $PCE = 2{,}576.10^{-2}$

TABLE 2.6 – Plans de corrélation et valeurs de PCE du JTC non-linéaire sans ordre zéro suivant différentes valeurs de coefficient de non-linéarité, pour une autocorrélation, une corrélation vraie (images différentes du même objet) et une corrélation fausse.

k	Autocorrélation	Corrélation vraie	Corrélation fausse
1	 $PCE = 4{,}678.10^{-5}$	 $PCE = 4{,}594.10^{-5}$	 $PCE = 4{,}718.10^{-5}$
0,8	 $PCE = 1{,}118.10^{-4}$	 $PCE = 1{,}042.10^{-4}$	 $PCE = 9{,}692.10^{-5}$
0,6	 $PCE = 1{,}106.10^{-3}$	 $PCE = 6{,}334^{-4}$	 $PCE = 4{,}126.10^{-4}$
0,4	 $PCE = 5{,}390.10^{-2}$	 $PCE = 1{,}332.10^{-2}$	 $PCE = 5{,}000.10^{-3}$
0,2	 $PCE = 6{,}784.10^{-1}$	 $PCE = 7{,}155.10^{-2}$	 $PCE = 2{,}176.10^{-2}$

naître un objet suivant diverses orientations : le filtre adapté augmente considérablement le nombre de fausses alarmes tandis que le filtre de phase pure réduit le nombre de non-détections fausses.

La multi-corrélation permet de s'affranchir de cette lacune en combinant plusieurs images de références suivant différentes orientations de l'objet nous intéressant, augmentant ainsi artificiellement la robustesse du filtre de phase pure. Deux approches principales de multi-corrélation ont été apportées : le filtre composite et le filtre composite segmenté. Le filtre composite est défini par une simple addition des images de référence. Cette simplicité permet une implantation facilitée de la multi-corrélation. Néanmoins, ce filtre devient peu pertinent pour une corrélation sur un nombre élevé d'images de références. En effet, l'addition d'images dans engendre une saturation des pixels. De plus, du fait de l'opération, effectuée dans le domaine spatial, le même pixel peut se voir attribuer différentes régions de l'objet suivant l'orientation d'observation. En ce qui concerne le filtre composite segmenté, une sélection de pixel est effectuée dans le domaine spectral, annulant le problème induit par la saturation et permettant d'éviter qu'un même pixel corrèle avec différentes régions de l'objet dans l'image cible.

Pour finir, le corrélateur de Vander Lugt étant particulièrement adapté à une optimisation du filtre de corrélation en amont de la phase de corrélation, il est ainsi aisé de définir plusieurs filtres composites simples ou segmentés pour le même objet. Cette approche permet de créer des filtres pour un grand nombre de situations sans pour autant observer de saturation du filtre tout en limitant le temps de calcul nécessaire. Cette capacité du corrélateur de Vander Lugt d'optimisation d'un filtre pour un objet et une application donnés rendent cette architecture très performante pour une reconnaissance d'objet.

Le corrélateur à spectre joint, quant-à-lui est caractérisé par l'absence de création d'un filtre de corrélation. L'image de référence et l'image cible sont positionnées directement sur un même plan d'entrée, à partir duquel la corrélation est réalisée. Cette propriété permet de réduire le nombre d'étapes pour la corrélation, en premier lieu le nombre de transformées de Fourier, étape la plus coûteuse en calculs. En contrepartie, cette unique transformée de Fourier est appliquée sur une matrice beaucoup plus volumineuse, sur une aire au minimum 16 fois supérieure. Le corrélateur de phase pure classique présente des pics d'inter-corrélation très larges et de faible intensité, engendrant une très forte robustesse et donc une faible capacité de discrimination.

Pour pallier à cela, l'introduction d'une fonction de non-linéarité dans le plan de Fourier permet d'influer sur l'intensité et la largeur des pics d'inter-corrélation et donc de jouer sur la robustesse et la discrimination, entre les deux extrêmes que sont le corrélateur classique et le corrélateur binaire, suivant l'application voulue. De plus, ce type de corrélateur dispose d'une relation simple entre la position relative des pics d'inter-corrélation et la position relative de l'image référence et de sa correspondante dans l'image cible. L'ensemble de ces propriétés rend donc cette famille de corrélateurs particulièrement adaptée à une application de détection et de suivi d'objet dans l'image. En effet, dans une application de suivi, la corrélation est effectuée entre une image cible au temps t et une image référence au temps $t-1$. Celles-ci sont donc relativement similaires et il est recommandé d'utiliser un compromis entre la robustesse et la discrimination. De plus il n'est nullement besoin de réaliser un nouveau filtre à chaque laps de temps, ce qui serait le cas avec un corrélateur de Vander Lugt, réduisant d'autant le traitement et donc le temps de calcul. Également, il est possible d'introduire plusieurs images de référence dans le plan d'entrée, permettant, avec un traitement adapté, le suivi simultané de plusieurs objets. En contrepartie, ce filtre est peu adapté à la reconnaissance d'objets, un filtre optimisé pour un objet donné étant beaucoup plus compliqué à réaliser.

Enfin, le plan de corrélation présente un puissant pic d'auto-corrélation, pouvant être supprimé en calculant puis retranchant séparément au spectre de l'image d'entrée les auto-corrélations des images cibles et références. Cependant cette étape engendre le calcul de deux transformées de Fourier supplémentaires, augmentant d'autant le temps de calcul.

Ainsi, le corrélateur de Vander Lugt permet l'optimisation de filtres à un objet donné, le rendant très performant pour une application de reconnaissance. Quant au corrélateur à spectre joint, il est à même de limiter les étapes de calcul. De plus la simplicité de la relation entre la position des pics d'inter-corrélation et de l'objet dans l'image cible et sa capacité d'effectuer un compromis entre robustesse et discrimination à l'aide d'une seule image cible engendre un corrélateur adapté à une application de suivi d'objet dans une séquence d'images.

2.7 Conclusion

Dans ce chapitre nous avons présenté les deux architectures principales de corrélation optique, le corrélateur de Vander Lugt et le corrélateur à spectre joint.

Dans un premier temps, nous avons présenté différentes métriques permettant la détection du pic de corrélation et l'établissement d'un seuil afin de mettre en place un système de reconnaissance. Une optimisation de la courbe ROC a été introduite, afin de rendre compte de façon plus efficace que la courbe ROC classique des capacités de discrimination du classifieur.

Le corrélateur de Vander Lugt a été introduit, ainsi que ses différentes optimisations, le filtre de phase pure, le filtre composite et le filtre composite segmenté. Les différentes approches basées sur le corrélateur de Vander Lugt ont été comparées, ainsi que les critères de segmentation disponibles dans le cas du filtre composite segmenté. Enfin les critères de détection du pic de corrélation ont été évalués.

Finalement, nous avons décrit le corrélateur à spectre joint ainsi que deux optimisations permettant d'améliorer les performances du filtre, à savoir annuler le risque de chevauchevement du pic d'autocorrélation et des pics d'intercorrélation ainsi que les capacités de discrimination et de robustesse du corrélateur.

Se basant sur les capacités de suivi et d'identification de la corrélation, nous présentons dans les chapitres suivant une application du VLC à l'identification (chapitre 3) et du JTC au suivi dans une séquence d'images (chapitre 4).

Deuxième partie

Corrélation pour l'identification et le suivi

Chapitre 3

Application de la corrélation pour l'identification

Sommaire

Dans ce chapitre est présentée une application de la corrélation de Fourier [86] à l'identification. Comme explicité précédemment au chapitre 2, l'identification par corrélation est une approche réalisée originellement optiquement. L'implantation optique, bien que potentiellement prometteuse en terme de temps de calcul, est limitée physiquement par les interfaces nécessaires à l'affichage des filtres de corrélation ou des plans d'entrée. De plus l'implantation optique ne rend accessible pour optimisation qu'une partie du processus de corrélation. La corrélation optique comporte deux architectures distinctes, le corrélateur de Vander Lugt et le corrélateur à spectre joint. Bien que tous deux permettent théoriquement une utilisation pour des applications d'identification et de suivi, leurs performances diffèrent sensiblement. Le VLC, par sa capacité d'optimisation de filtres de corrélation à l'objet et l'application voulue, ainsi que la très grande discrimination du filtre de phase pure, est le plus adapté à l'identification.

La problématique majeure d'un système d'identification est sa capacité à différencier deux objets distincts tout en permettant la reconnaissance d'un même objet malgré l'existence de transformations. Ainsi, différentes optimisations du corrélateur ont été proposées [86, 87]. Celles-ci n'ont généralement concerné que certaines étapes du processus de corrélation, à savoir les filtres de corrélation, l'image cible, le plan d'entrée et le plan de Fourier. L'apparition d'unités de calcul performantes au cours des deux dernières décennies ont rendu possible une implantation entièrement numérique de la corrélation, ouvrant la voie à de nouvelles optimisation des corrélateurs. Néanmoins, peu de travaux ont été menés sur le plan de corrélation en lui-même. Le travail présenté dans cette partie propose l'utilisation d'une méthode de débruitage au plan de corrélation à l'aide d'un modèle linéaire [88]. Pour ce faire, une série de pics de corrélation a été modélisée, et différents bruits de corrélation ont été générés et incorporés au modèle.

Nous commençons dans ce chapitre par présenter les étapes de réalisation du modèle linéaire, son application pour la décomposition du plan de corrélation et son débruitage. Nous voyons ensuite une étape de centrage du pic de corrélation, indispensable à l'utilisation du modèle linéaire, et une comparaison des fonctions de modélisation du signal. Puis nous évaluons les performances de notre méthode par rapport à la corrélation sans débruitage.

3.1 Le modèle linéaire pour le débruitage du plan de corrélation

Le corrélateur de Vander Lugt permet d'obtenir, à partir d'une comparaison d'une image référence avec une image cible, un plan de corrélation comprenant un pic dont l'intensité est dépendant de la similarité entre ces images. Malheureusement, le plan de corrélation présente, lorsque ces dernières ne sont pas identiques, un fort niveau de bruit. Un exemple de corrélation à l'aide de l'architecture Vander Lugt est présentée en figure 3.1. L'image référence utilisée comprend une personne de face (Fig. 3.1c). Deux plans sont présentés, obtenus à l'aide d'un filtre de phase pure. L'un utilise une image cible (Fig. 3.1a) identique à l'image référence (Fig. 3.1d), l'autre plan (Fig. 3.1e) utilise une image cible comprenant le visage de profil de la personne (Fig. 3.1b). Il s'agit donc de deux images différentes du même objet, prises dans les mêmes conditions de luminosité. Nous observons en figure 3.1d un plan d'auto-corrélation (images référence et cible identique) présentant un pic de corrélation intense et fin. L'autre plan (Fig. 3.1e), quant-à-lui, présente bien un pic de corrélation, mais il est relativement faible comparativement au plan obtenu et le plan comprend surtout un fort niveau de bruit de corrélation. Ce fort niveau de bruit est problématique car il engendre une plus forte probabilité de non détection fausse, deux images de la même personne risquant d'engendrer un PCE plus faible que lors de la corrélation de deux personnes différentes.

Pour remédier à cela et améliorer les performances du corrélateur, différentes approches ont été proposées (e.g. augmentation du nombre de références avec l'utilisation d'un filtre composite, optimisation du critère de décision). Nous introduisons ici une nouvelle étape, permettant de réduire le niveau de bruit du plan de corrélation avant utilisation d'un critère de détection du pic de corrélation. L'idée est d'isoler le pic de corrélation, toujours présent sur le plan, du bruit de corrélation. Pour ce faire, nous proposons l'utilisation d'un modèle linéaire (LM) [120, 121], permettant une décomposition du signal. Dans cette partie, nous commençons par

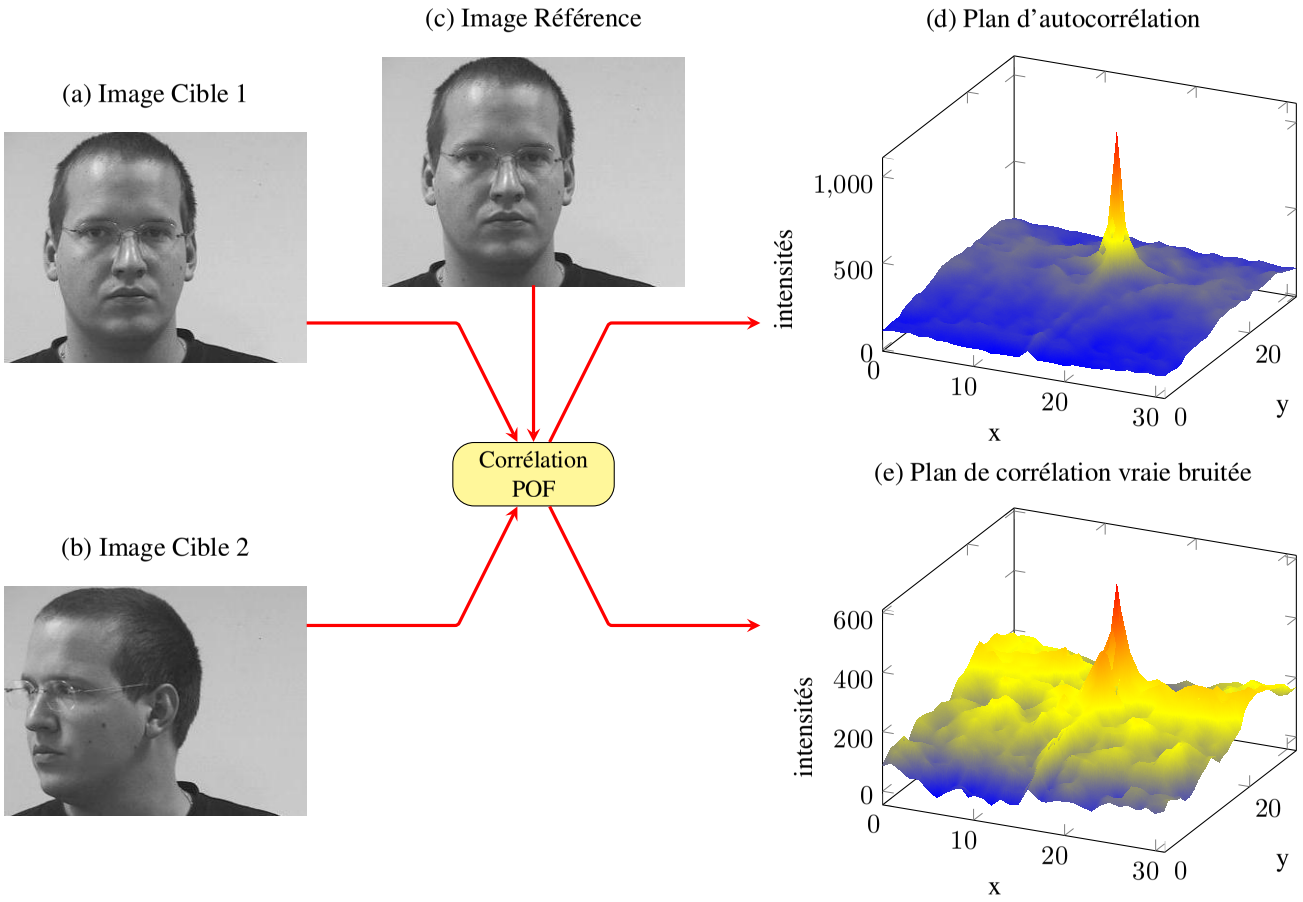


FIGURE 3.1 – Exemples de plans de corrélation obtenus avec le filtre POF : (a) visage de face ; (b) visage de profil ; (c) image référence ; (d) plan de corrélation obtenu en effectuant la corrélation de l'image (a) avec l'image référence ; (d) plan de corrélation obtenu en effectuant la corrélation de l'image (b) avec l'image référence.

introduire le modèle linéaire pour des images, puis nous détaillons le choix des différents régresseurs composant notre modèle, c'est-à-dire les caractéristiques statistiques permettant la décomposition du signal. Enfin nous présentons l'étape de reconstruction du plan de corrélation.

3.1.1 Décomposition en modèle linéaire

L'utilisation d'un modèle linéaire consiste en une décomposition du signal observé en une somme finie de régresseurs pondérés [120]. Cette méthode est notamment utilisée pour décomposer des signaux temporels en imagerie par résonance magnétique fonctionnelle (IRMf). Également, Reynaud et al. [121] utilisent un modèle linéaire pour débruiter de séquences vidéo obtenues par imagerie optique cérébrale, la série temporelle de chaque pixel de la séquence étant traitée séparément.

Cette méthode nécessite, pour être utilisée, d'être en mesure de décrire la forme générale du signal. Dans le cadre de la corrélation, la forme générale du signal est également connue : il s'agit d'un pic de corrélation centré dans le plan de corrélation. L'évasement et la puissance du pic de corrélation est dépendante de la ressemblance entre les images cible et référence. Le principe générale d'une telle approche est de décomposer le signal en une combinaison linaire. Le problème central est donc la caractérisation précise des régresseurs, c'est à dire des caractéristiques de décomposition du signal.

La décomposition du plan de corrélation P_c en une combinaison linéaire de régresseurs Y_i est donnée par l'équation 3.1. M est un nombre entier donné, β_i correspond à la pondération du régresseur Y_i et R au bruit résiduel, c'est-à-dire la partie du plan de corrélation inexpliquée par le modèle linéaire et dans lequel une portion du signal ou du bruit peut être encore présente.

$$P_C = \sum_{i=1}^{M} \beta_i Y_i + R \tag{3.1}$$

En pratique, il est commode de réécrire l'équation 3.1 sous forme matricielle :

$$P_C = \mathbf{Y}\beta + R, \tag{3.2}$$

où $\mathbf{Y}$ est le vecteur colonne contenant les différents régresseurs Y_i et β est le vecteur ligne contenant les poids correspondant aux régresseurs. En supposant blanc les bruit résiduels, le meilleur estimateur non biaisé β^+ de β, où $\beta^+ = \mathbf{Y}^+ Pc$ est obtenu à l'aide de la matrice pseudo-inverse $\mathbf{Y}^+$ de $\mathbf{Y}$, c'est-à-dire : $\mathbf{Y}^+ = (\mathbf{Y}^t\mathbf{Y})^{-1}\mathbf{Y}^t$.

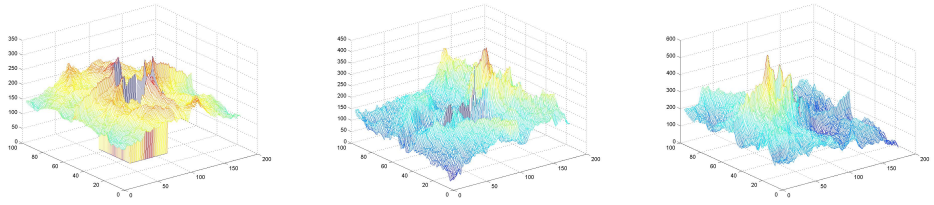
3.1.2 Choix des régresseurs

Afin de décomposer efficacement le signal, il est nécessaire de définir des régresseurs correspondant à l'application souhaitée. Nous recherchons ici à être en mesure d'isoler le pic de corrélation du bruit présent sur le plan de corrélation. Pour ce faire nous devons donc définir deux modèles : (i) un ensemble de régresseurs représentant le bruit du plan de corrélation ; (ii) un ensemble de régresseurs représentant le signal, c'est à dire le pic de corrélation. Ces modèles doivent être le plus généraux possibles. En particulier, le modèle du bruit doit être en mesure d'expliquer le bruit présent dans le plan lorsqu'il y a un pic de corrélation et lorsque ce pic est absent du plan.

3.1.2.1 Modélisation du bruit

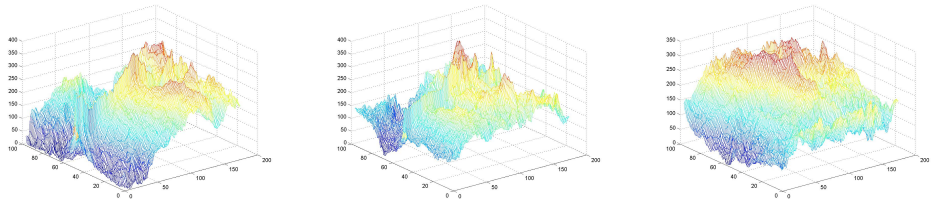
La première étape de décomposition du plan de corrélation consiste à modéliser le bruit présent dans le plan de corrélation, c'est-à-dire, tout ce qui n'est pas le pic de corrélation. Il s'agit donc de définir un modèle représentant le plan de corrélation excepté le pic central lorsqu'il y a présence d'un pic et l'ensemble du plan de corrélation en l'absence de pic de corrélation. À défaut d'une modélisation efficace du bruit de corrélation,

celui-ci étant extrêmement dépendant du fond de l'image et du système d'acquisition, nous avons choisi de créer une base de plans de corrélation réalisés à partir du filtre utilisé pour l'étape d'identification. Les figures 3.2 et 3.3 représentent des exemples de plans utilisés pour la modélisation du bruit. Deux types de bruit peuvent être distingués, à savoir le bruit de corrélation, c'est-à-dire le bruit présent autour du pic de corrélation, et le bruit de non-corrélation, défini par l'ensemble du plan en incluant le pic de corrélation, lorsque l'image cible et référence ne sont pas issues du même objet. En effet, un pic de corrélation important peut être présent lors de la comparaison d'images issues d'objets différents, compromettant la capacité de discrimination du filtre. De tels phénomènes doivent donc être pris en compte lors de la caractérisation des régresseurs du bruit. La figure 3.2 contient les plans réalisés avec des images cibles contenant l'objet à identifier (celui utilisé en référence) permettant la génération du bruit de corrélation et la figure 3.3 les plans réalisés avec des images cibles ne contenant pas l'objet utilisé pour l'image de référence, correspondant au bruit de non-corrélation. Pour le premier cas, le pic de corrélation a été retiré des plans de corrélation. Une zone de 20 pixels autour du centre du plan de corrélation a été supprimée.



(a) (b) (c)

FIGURE 3.2 – Exemples de plans de corrélation utilisés pour la réalisation du modèle du bruit pour le cas où l'objet présent dans l'image de référence est présent dans l'image cible. Une région de 20 pixels autour du centre du plan a été supprimée pour retirer le pic de corrélation.



(a) (b) (c)

FIGURE 3.3 – Exemples de plans de corrélation utilisés pour la réalisation du modèle du bruit pour le cas où l'objet présent dans l'image de référence n'est pas présent dans l'image cible.

Finalement, en utilisant cette méthode de sélection des régresseurs, nous obtenons le vecteur colonne $\mathbf{Y^{bruit}}$, donné par l'équation 3.3 avec $bruit_1 \cdots bruit_n$ représentant les n plans de corrélation utilisés.

$$\mathbf{Y^{bruit}} = \begin{bmatrix} bruit_1 \\ \vdots \\ bruit_n \end{bmatrix} \tag{3.3}$$

Le bruit compris dans le plan de corrélation est dépendant du filtre utilisé pour la corrélation et de la base de donnée sur laquelle il est appliqué. L'étape de sélection des régresseurs du bruit est donc nécessairement adaptée à chaque expérimentation et application.

3.1.2.2 Modélisation du signal

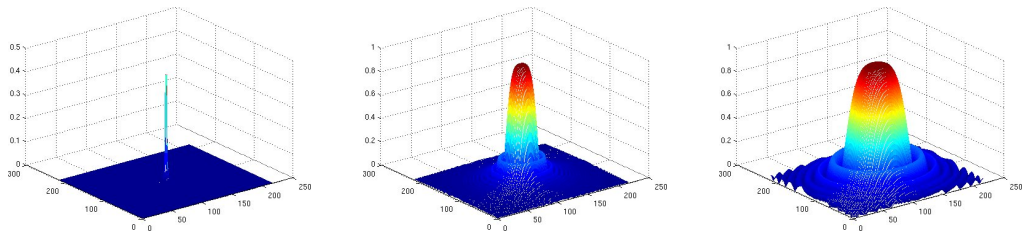
Le débruitage du plan de corrélation consiste à retenir du plan seulement l'information désirée, c'est-à-dire le pic de corrélation ($\mathbf{Y^{pic}}$) et à retirer le bruit engendré par la corrélation ($\mathbf{Y^{bruit}}$). L'étape la plus importante est donc celle de la modélisation du pic de corrélation, dont la forme générale est connue. Pour ce faire, et en se basant sur les données obtenues après application d'un filtre POF, nous avons défini deux fonctions de modélisation du pic de corrélation, qui seront comparées en partie 3.2.2, page 71. Nous avons tout d'abord expérimenté une approximation du pic de corrélation à l'aide d'une fonction sinus cardinal tridimensionnelle. En posant $sin_{2D}(i,j)$ le sinus cardinal, avec (i,j) le pixel correspondant dans le plan de sortie, i_0, j_0 les moyennes et σ_i, σ_j représentent les écarts-type en i et j, le sinus cardinal tridimensionnel est donné par :

$$sinc_{2D}(i,j) = \left| \frac{\sin(\frac{(i-i_0)^2}{2\sigma_i^2})}{\frac{(i-i_0)^2}{2\sigma_i^2}} \times \frac{\sin(\frac{(j-j_0)^2}{2\sigma_j^2})}{\frac{(j-j_0)^2}{2\sigma_j^2}} \right| \tag{3.4}$$

De la même façon, une seconde approximation du pic de corrélation a été expérimentée, basée cette fois ci sur une fonction inverse tridimensionnelle, $inv_{2D}(i,j)$, définie de la façon suivante :

$$inv_{2D}(i,j) = \frac{1}{i - i_0} \times \frac{1}{j - j_0} \tag{3.5}$$

Enfin, afin d'être en mesure d'expliquer la variété des pics de corrélation il est nécessaire de créer une kyrielle de pics modélisés avec différentes largeurs. Les figures 3.4 et 3.5 présentent des exemples de cette base de pics créés pour la fonction sinus cardinal et inverse, respectivement, en prenant $i_0 = j_0 = 1$ (Fig. 3.4a et 3.5a), $i_0 = j_0 = 50$ (Fig. 3.4b et 3.5b) et $i_0 = j_0 = 100$ (Fig. 3.4c et 3.5c).

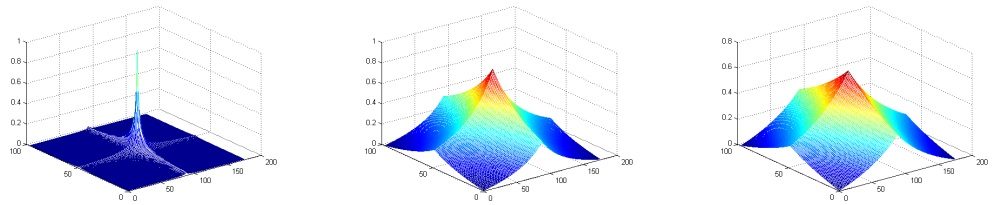


(a) (b) (c)

FIGURE 3.4 – Modélisation du pic de corrélation avec un sinus cardinal, avec $i_0 = j_0 = 1$ (Fig. 3.4a), $i_0 = j_0 = 50$ (Fig. 3.4b) et $i_0 = j_0 = 100$ (Fig. 3.4c).

L'ensemble de ces pics de corrélation modélisés sont finalement concaténés dans un vecteur colonne Y^{pic}, avec $pic_1 \cdots pic_n$ représentant les n plans de corrélation modélisés utilisés, soit avec la fonction sinus cardinal, soit avec la fonction inverse :

$$\mathbf{Y^{pic}} = \begin{bmatrix} pic_1 \\ \vdots \\ pic_n \end{bmatrix} \tag{3.6}$$



(a) (b) (c)

FIGURE 3.5 – Modélisation du pic de corrélation avec une fonction inverse, avec $i_0 = j_0 = 1$ (Fig. 3.5a), $i_0 = j_0 = 50$ (Fig. 3.5b) et $i_0 = j_0 = 100$ (Fig. 3.5c)

La sélection de la fonction permettant la modélisation du pic de corrélation est une étape importante car de celle-ci découle la capacité d'isolement du pic dans un plan de corrélation réel. Le choix de la fonction de modélisation et le nombre de régresseurs utilisés sera donc discuté par la suite afin d'obtenir un modèle robuste.

3.1.3 Création du modèle linéaire

À l'aide du vecteur colonne $\mathbf{Y^{pic}}$ correspondant à la modélisation des différents bruits et $\mathbf{Y^{bruit}}$ correspondant à la modélisation de la réponse on crée leur concaténation, le vecteur colonne $\mathbf{Y}$, représentant la modélisation du signal complet, en faisant l'omission du bruit résiduel :

$$\mathbf{Y} = \begin{bmatrix} Y^{pic} \\ Y^{bruit} \end{bmatrix}. \tag{3.7}$$

3.1.4 Débruitage du plan de corrélation

3.1.4.1 Principe

Etant en possession d'un modèle permettant d'expliquer le signal et le bruit présents dans le plan de corrélation, nous sommes maintenant en mesure de réaliser une décomposition linéaire du plan de corrélation. L'étape de décomposition linéaire du plan est définie par $Pc^{'} = \mathbf{Y}\beta^{+}$, $Pc^{'}$ étant le plan de corrélation reconstruit à l'aide du modèle, et avec $\beta^{+} = \mathbf{Y}^{+}Pc$. La décomposition en modèle linéaire est résumée en figure 3.6 : le plan de corrélation vraie ou fausse (P_c) est la combinaison linaire des régresseurs du bruit $\mathbf{Y^{bruit}}$ et des régresseurs du signal $\mathbf{Y^{signal}}$, pondérés respectivement par β_1 et β_2, ainsi qu'un bruit résiduel R.

Une fois le plan de corrélation Pc' décomposé à l'aide du modèle linéaire, une partie de celui-ci peut rester inexpliquée par les régresseurs choisis. Ces bruits résiduels R se calculent de la façon suivante :

$$R = Pc - Pc'. \tag{3.8}$$

Pour finalement reconstruire un plan de corrélation optimisé Pc^{opt} en omettant les régresseurs du bruit $\mathbf{Y^{bruit}}$:

$$Pc^{opt} = \mathbf{Y^{pic}}\beta^{+}_{pic} + R. \tag{3.9}$$

Le bruit résiduel est conservé dans le plan de corrélation optimisé afin de prendre en compte les éventuels signaux et bruits non explicables par notre modèle linéaire. Le plan de corrélation obtenu par application de notre modèle linéaire est finalement une combinaison linaire des régresseurs du signal, auquel on a ajouté une partie du bruit, le bruit résiduel. Le bruit du plan de corrélation explicable à l'aide de notre modèle a été supprimé du plan de corrélation final, appelé "plan de corrélation".

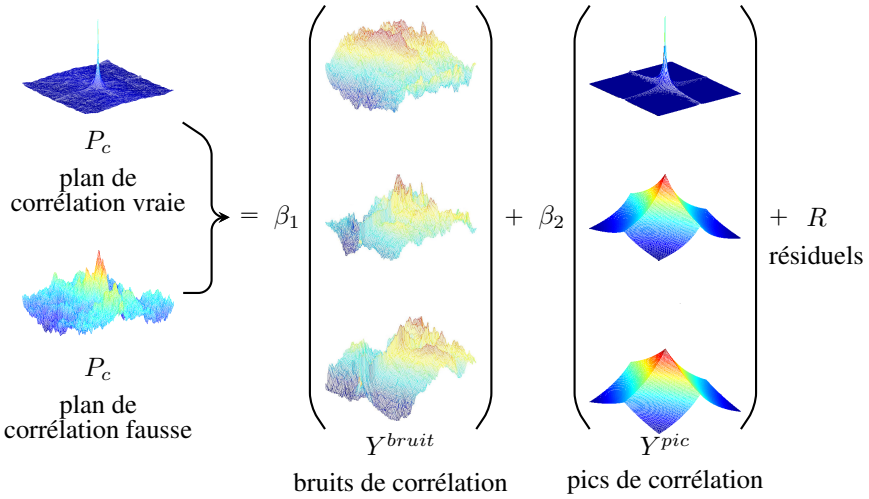


FIGURE 3.6 – Processus de décomposition du plan de corrélation à l'aide du modèle linéaire.

3.1.4.2 Expérimentation

Nous illustrons ici notre méthode de débruitage à l'aide des fonction sinus cardinal et inverse tridimensionnelles. Nous utilisons un filtre POF dont les images cible et référence sont présentées en figure 3.7. L'image référence utilisée pour la fabrication du filtre (Fig. 3.7a) est la personne 1, image 20 (visage de face) de la base PHPID [116]. En ce qui concerne l'image cible, il s'agit, pour la corrélation vraie, de l'image 20 de la personne 1 (Fig. 3.7a) et de l'image 20 de la personne 2 pour la corrélation fausse (Fig. 3.7b).

(a) (b)

FIGURE 3.7 – Images référence et cible utilisées pour la corrélation avec le filtre POF : (a) images de référence ; (b) image cible.

Le bruit a été créé à l'aide des plans de corrélation obtenus après application du filtre sur les personnes 10 à 15 de la base PHPID ainsi que sur les 5 plans retournant le plus faible PCE de la personne 1, le pic central ayant été supprimé.

Les figures 3.8 et 3.9 présentent les résultats de la décomposition du plan de corrélation à l'aide de la fonction sinus cardinal tridimensionnelle pour le cas d'une corrélation vraie et fausse, respectivement. Nous représentons les plans de corrélation originaux (Fig. 3.8a et 3.9a) et après débruitage (Fig. 3.8b et 3.9b), ainsi que les signaux (Fig. 3.8c et 3.9c), bruits (Fig. 3.8d et 3.9d) et résiduels (Fig. 3.8e et 3.9e). De la même façon, la décomposition à l'aide de la fonction inverse tridimensionnelle pour le cas d'une corrélation vraie et fausse

est représentée par les figures 3.10 et 3.11. Sont également représentés les plans de corrélation originaux (Fig. 3.10a et 3.11a) et après débruitage (Fig. 3.10b et 3.11b), les signaux (Fig. 3.10c et 3.11c), bruits (Fig. 3.10d et 3.11d) et résiduels (Fig. 3.10e et 3.11e).

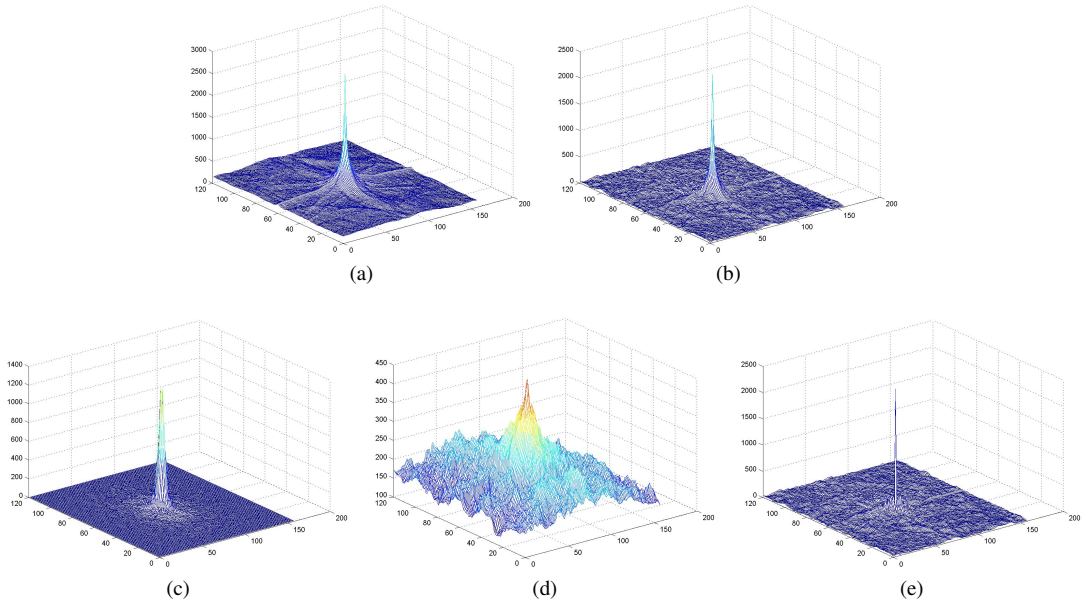


FIGURE 3.8 – Décomposition du plan de corrélation lorsque les images référence (Fig. 3.7a) et cible (Fig. 3.7b) sont identiques avec utilisation de la fonction sinus cardinal tridimensionnelle : (a) plan de corrélation ; (b) plan reconstruit ; (c) signal ; (d) bruit ; (e) résiduels.

Nous observons en premier lieu que le pic de corrélation a été précisément reconstitué pour le cas de la corrélation vraie pour les deux fonctions de modélisation du signal (Fig. 3.8c et 3.10c). À l'inverse, celui-ci est faible et évasé pour le cas de la corrélation fausse (Fig. 3.9c 3.11c). Les composants du bruit, quant à eux, sont très peu puissants, avec un maximum de $400a.u.$[1] et $250a.u.$ pour la corrélation vraie (fonction sinus cardinal et inverse, respectivement), avec des bruits résiduels présentant un pic de $1500a.u.$ et $400a.u.$. La grande puissance de signal encore contenue dans le bruit résiduel est due à une lacune dans l'explication d'un des deux composants du modèle (le bruit et le signal). Le modèle du bruit ne pouvant être créé que expérimentalement, il est donc plus à même de ne pas expliquer efficacement le bruit de corrélation. Enfin, le plan de corrélation reconstruit présente bien un pic net pour le cas de la corrélation vraie (Fig. 3.10b). À l'inverse, le pic de corrélation a été totalement supprimé pour le cas de la corrélation fausse (Fig. 3.11b).

Notre méthode de débruitage permet donc d'accentuer la séparation des cas de vraie et de fausse corrélation et donc de réaliser un débruitage du plan de corrélation retourné par un filtre VLC. Comme nous venons de le voir, différents paramètres influent sur les performances de la décomposition en modèle linéaire, notamment le nombre et les choix des régresseurs introduits dans le modèle. En effet, le choix des régresseurs étant une étape primordiale de notre approche, il est nécessaire d'étudier leurs effets afin d'optimiser la décomposition et donc la classification.

1. $a.u.$: "absolute unity"

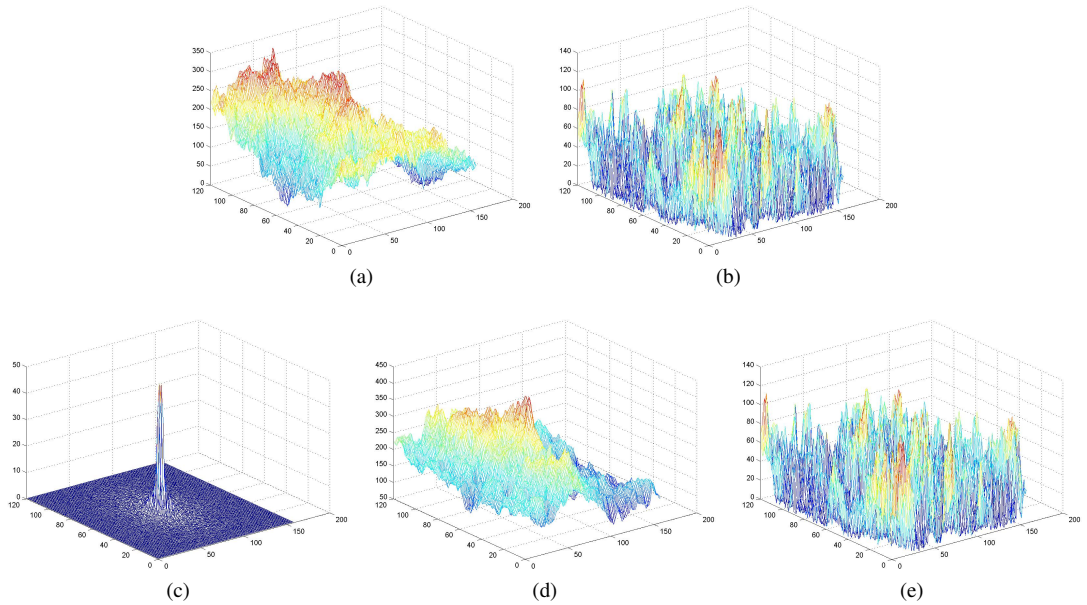


FIGURE 3.9 – Décomposition du plan de corrélation lors de non correspondance entre les images référence (Fig. 3.7a) et cible (Fig. 3.7b) avec utilisation de la fonction sinus cardinal tridimensionnelle : (a) plan de corrélation ; (b) plan reconstruit ; (c) signal ; (d) bruit ; (e) résiduels.

3.2 Les paramètres de l'identification

La pertinence de notre méthode de débruitage par décomposition en modèle linéaire est conditionnée par différents paramètres. Dans le but d'obtenir une classification efficace nous étudions dans cette partie l'effet des paramètres accessibles : le pré-traitement de l'image cible, le centrage du pic de corrélation, le choix de la fonction de modélisation du signal 3.2.2 et des régresseurs du bruit.

L'étude de notre approche a été réalisée sur la base PHPID. Celle-ci est constituée de 15 personnes différentes, et de 93 images par personnes. 39 images par personnes ont été utilisées dans cette étude, présentant des orientations du visage allant de $-90°$ à $+90°$ avec un pas de $15°$ (la position à $0\circ$ correspondant à une personne de face) dans la direction horizontale et sont de $-10°$, $0°$ et $+10°$ dans la direction verticale. Une version réduite de la base, ne comportant que 5 personnes, a également été utilisée.

3.2.1 Centrage du pic de corrélation

Notre approche utilise une modélisation d'un signal centré sur le plan de corrélation. Ainsi, elle présuppose un pic de corrélation apparaissant au centre du plan. Or la position de celui-ci est dépendante de la position des régions corrélant sur les images cible et référence. Il est donc nécessaire d'effectuer un prétraitement du plan de corrélation afin d'obtenir un pic de corrélation localisé au centre du plan.

Notre méthode de centrage du pic de corrélation est présentée en figure 3.12. Une observation des plans de corrélation nous permet de définir une région d'apparition du pic de corrélation de $174 \times 119\ px$ centrée dans un plan de $314 \times 238\ px$ (taille des images utilisées), correspondant à la moitié des dimensions du plan original. Tout maximum local apparaissant en dehors de cette zone n'est pas considéré comme un pic de corrélation. Une telle zone nous permet finalement dans tous les cas de définir un redécoupage du plan de corrélation de $70 \times 70\ px$ autour du pic de corrélation détecté.

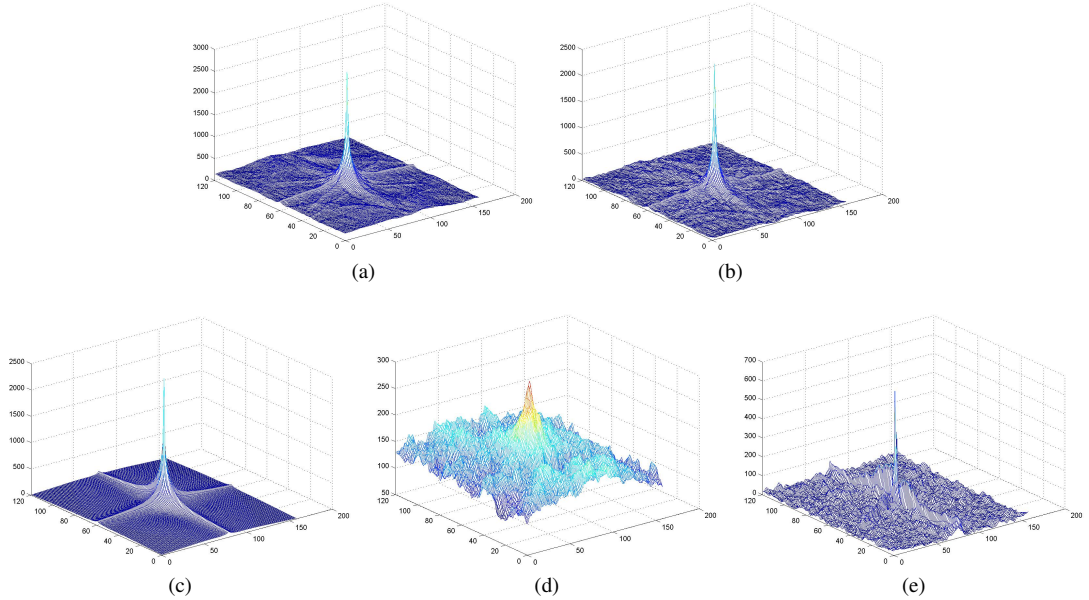


FIGURE 3.10 – Décomposition du plan de corrélation lorsque les images référence (Fig. 3.7a) et cible (Fig. 3.7b) sont identiques avec utilisation de la fonction inverse tridimensionnelle : (a) plan de corrélation ; (b) plan reconstruit ; (c) signal ; (d) bruit ; (e) résiduels.

La figure 3.13 illustre un exemple de l'étape de centrage du pic de corrélation. Le plan de corrélation original 3.13a contient un pic de corrélation non centré sur l'image, positionné aux coordonnées $(201{,}120)$ sur un plan de taille $314 \times 238px$. Nous observons en figure 3.13b le plan final, qui a été rogné autour du pic de corrélation de façon à le retrouver au centre du plan de corrélation. Le pic est positionné aux coordonnées $(79{,}60)$ sur un plan de $158 \times 120px$. Il est donc désormais possible d'utiliser une fonction de modélisation centrée sur le plan de corrélation, permettant une décomposition du signal.

3.2.2 Effet de la fonction utilisée pour la modélisation du signal

Nous avons défini deux modélisations différentes du signal en partie 3.1.2.2, page 66. Afin de déterminer celle qui s'approche le plus du signal réel, nous avons expérimenté notre méthode de débruitage sur une auto-corrélation, une fausse corrélation et une vraie corrélation à l'aide d'une image cible présentant la même personne que celle contenue dans l'image référence mais dans une position différente. Nous utilisons ici un filtre POF avec la personne 1, image 20 (personne de face), de la base PHPID. Les images cible et référence sont présentées en figure 3.14. La corrélation a été calculée avec comme image cible, l'image 20 de la personne 1 (Fig. 3.14a pour l'auto-corrélation, l'image 20 de la personne 2 pour la corrélation fausse (Fig 3.14b) et l'image 25 de la personne 1 pour la corrélation vraie avec changement d'orientation du visage (Fig 3.14c). Nous présentons les résultats avec une modélisation du signal utilisant la fonction sinus cardinal (Fig. 3.15, 3.16 et 3.17) et la fonction inverse (Fig. 3.18, 3.19 et 3.20). Le bruit a été créé à l'aide des plans de corrélation obtenus après application du filtre sur les personnes 10 à 15 de la base PHPID ainsi que sur les 5 plans retournant le plus faible PCE de la personne 1, une zone de $30 \times 30px$ de côté au centre du plan ayant été supprimé afin d'annuler le plan de corrélation. Sont représentés le plan de corrélation original et après débruitage, ainsi que les signaux, bruits et résiduels.

Nous présentons le plan de corrélation original (Fig. 3.15a à 3.20a) et optimisé (Fig. 3.15b à 3.20b) ainsi

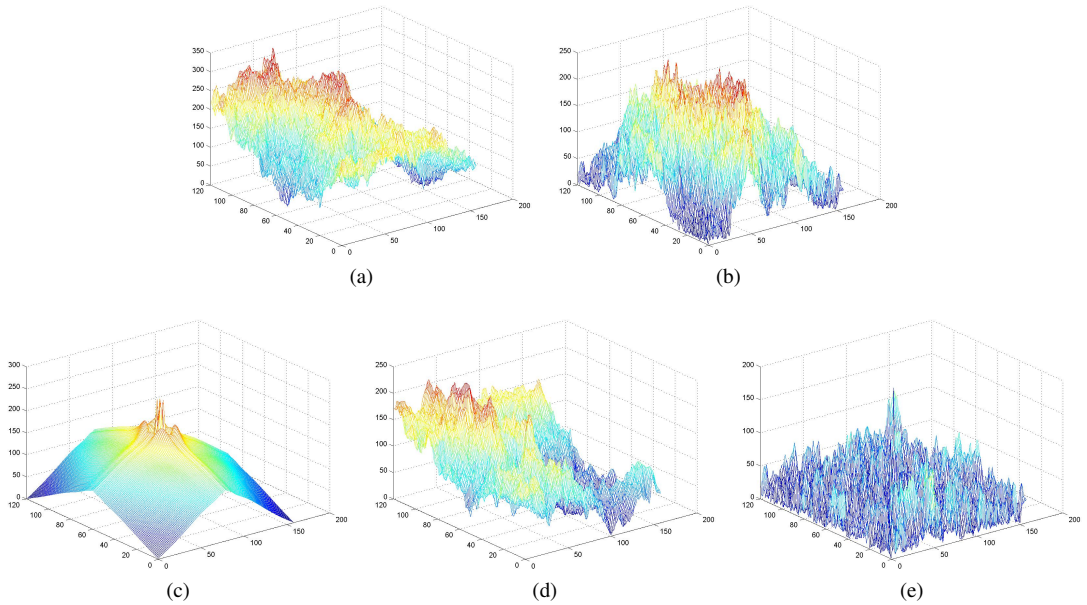


FIGURE 3.11 – Décomposition du plan de corrélation lors de non correspondance entre les images référence (Fig. 3.7a) et cible (Fig. 3.7b) avec utilisation de la fonction inverse tridimensionnelle : (a) plan de corrélation ; (b) plan reconstruit ; (c) signal ; (d) bruit ; (e) résiduels.

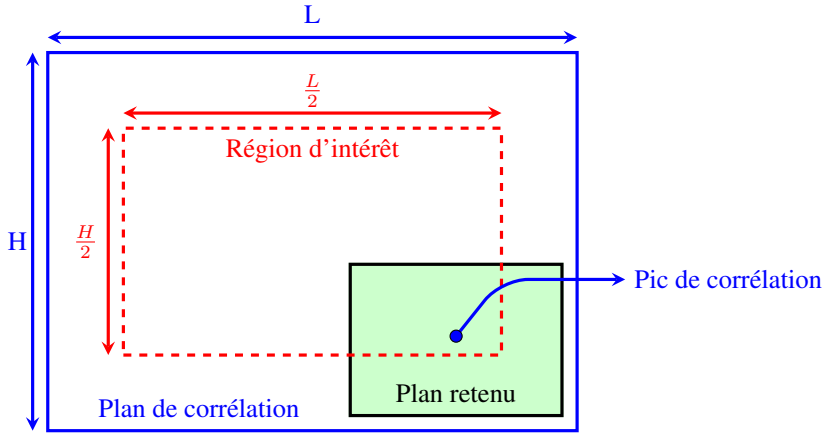
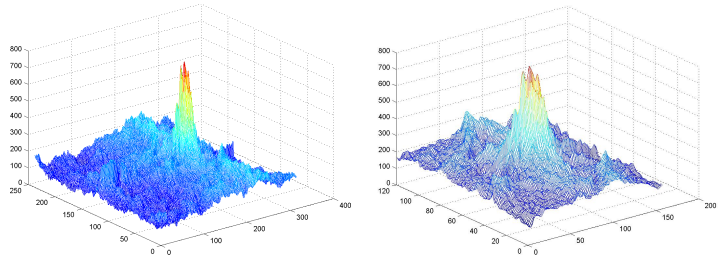


FIGURE 3.12 – Principe du centrage du pic de corrélation. H et L représentent la taille du plan de corrélation.

que les régresseurs du signal (Fig. 3.15c à 3.20c). Tout d'abord, nous n'observons pas de différence notable dans la reconstruction du plan entre les deux méthodes de modélisation pour ce qui est de l'auto-corrélation (Fig. 3.15b et 3.18b). En ce qui concerne la corrélation fausse, nous pouvons remarquer que le pic central de corrélation a été supprimé dans les deux cas (Fig. 3.18c et 3.15c). Bien que les composants du signal sont plus proches de la réalité avec la fonction inverse (Fig. 3.18c) comparativement à une utilisation de la fonction sinus cardinal (Fig. 3.15c), la fonction sinus cardinal engendre un plan de corrélation composé uniquement de résiduel alors que le bruit est encore très présent lors de l'utilisation de la fonction inverse. Finalement,

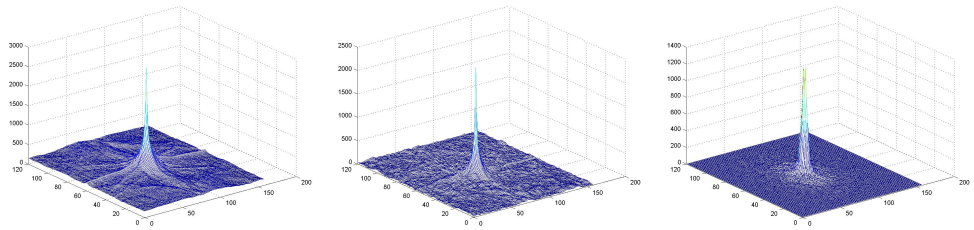


(a) (b)

FIGURE 3.13 – Centrage du pic de corrélation : (a) plan original ; (b) plan centré sur le pic de corrélation.

(a) (b) (c)

FIGURE 3.14 – Images référence et cible utilisées pour la corrélation avec le filtre POF : (a) image de référence ; (c) image cible pour la corrélation fausse ; (c) image cible pour la corrélation vraie bruitée.



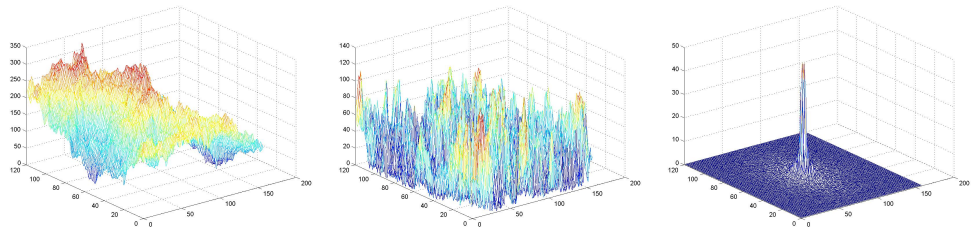
(a) (b) (c)

FIGURE 3.15 – Décomposition du plan de corrélation avec la fonction sinus cardinal lors de correspondance entre le filtre et l'image de référence : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).

lors de l'utilisation d'une image cible présentant la même personne que sur l'image référence mais avec un changement d'orientation (Fig. 3.20c et 3.17c) on observe un pic plus intense avec la fonction inverse dans le signal reconstruit.

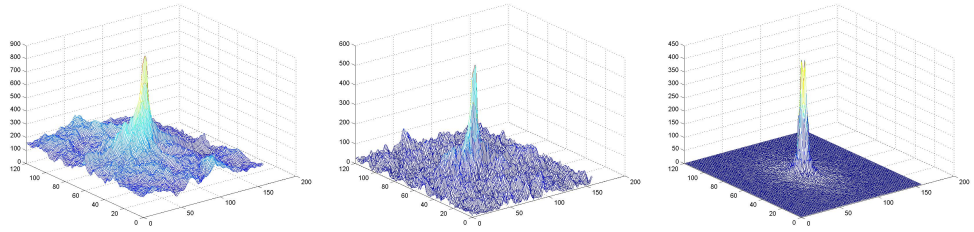
Afin de départager les fonctions de modélisation du signal, nous présentons en figure 3.21 les courbes ROC obtenues sur l'ensemble de la base PHPID pour la fonction inverse (Fig. 3.21a) et sinus cardinal (Fig. 3.21b) tridimensionnelles. Les résultats ont été obtenus en utilisant 10 signaux de modélisation. Nous observons des résultats largement supérieurs pour le cas de la fonction sinus cardinal, avec une AUC [2] de $0{,}745$ pour la fonction inverse et de $0{,}973$ pour la fonction sinus cardinal. De plus nous observons un taux de vrai positifs de

2. Area Under Curve : Aire sous la courbe



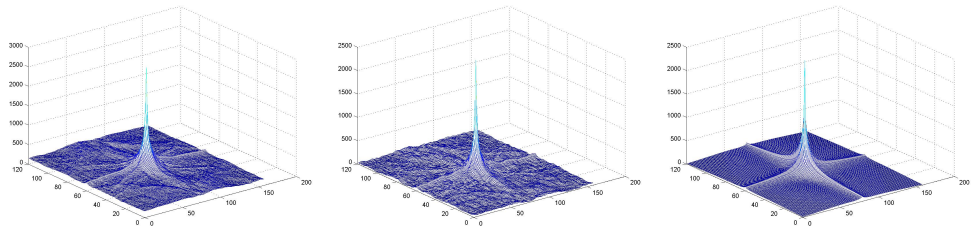
(a) (b) (c)

FIGURE 3.16 – Décomposition du plan de corrélation avec la fonction sinus cardinal lors de non correspondance entre le filtre et l'image de référence : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).



(a) (b) (c)

FIGURE 3.17 – Décomposition du plan de corrélation avec la fonction sinus cardinal avec changement d'orientation du visage dans l'image cible : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).

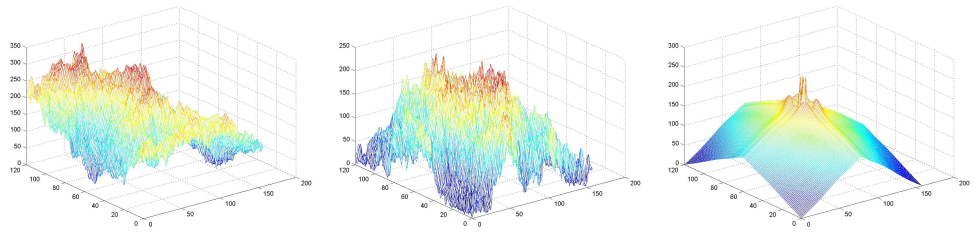


(a) (b) (c)

FIGURE 3.18 – Décomposition du plan de corrélation avec la fonction inverse lors de correspondance entre le filtre et l'image de référence : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).

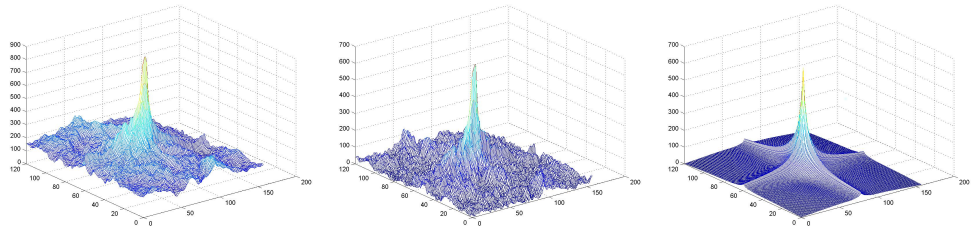
$20{,}5\%$ pour un taux de faux négatifs de 0% pour la fonction inverse alors qu'il est de $61{,}54\%$ pour la fonction sinus cardinal. Finalement, pour un $TPR = 1$ on a un FPR de $3{,}1\%$ pour la fonction inverse et de 16.5% pour la fonction sinus cardinal.

La décomposition en modèle linaire est donc plus efficace lors d'une utilisation d'un modèle du signal composé d'une série de fonctions sinus cardinal tridimensionnelles que lors d'une modélisation avec une fonction inverse, bien qu'étant moins proche visuellement du signal réel.



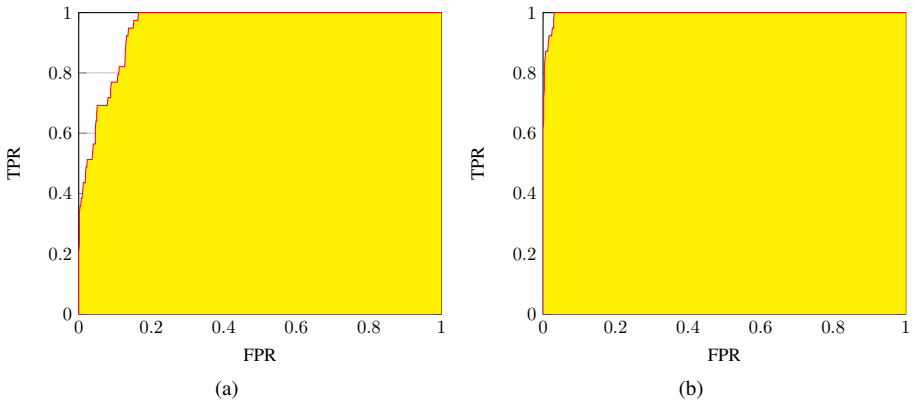
(a) (b) (c)

FIGURE 3.19 – Décomposition du plan de corrélation avec la fonction inverse lors de non correspondance entre le filtre et l'image de référence : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).



(a) (b) (c)

FIGURE 3.20 – Décomposition du plan de corrélation avec la fonction inverse avec changement d'orientation du visage dans l'image cible : (a) plan de corrélation ; (b) plan reconstruit ; signal (c).



(a) (b)

FIGURE 3.21 – Courbes ROC pour une corrélation sur l'ensemble de la base PHPID avec 10 signaux de modélisation du signal : (a) signal modélisé à l'aide de la fonction inverse tridimensionnelle ; (b) signal modélisé à l'aide de la fonction sinus cardinal tridimensionnelle.

3.2.3 Effet du nombre de signaux modélisés

Afin de déterminer le nombre optimum de signaux utilisés pour la modélisation du signal, nous avons généré les matrices de confusion (définies en partie 2.2.2, page 33) pour chacune des deux fonctions de modélisation du signal, sinus cardinal et inverse, avec des séries comprenant de 1 à 200 signaux. Les résultats ont

été calculés sur l'ensemble de la base PHPID en utilisant l'image 20 (visage de face) en référence d'un filtre de phase pure. L'effet du nombre de signaux est illustré en figure 3.22. L'évolution de l'aire sous la courbe est exprimée en fonction du nombre de signaux utilisés pour les deux fonctions. À titre indicatif, la valeur de l'aire sous la courbe pour la corrélation sans débruitage du plan de corrélation a été apportée au graphique ($AUC = 0{,}887$).

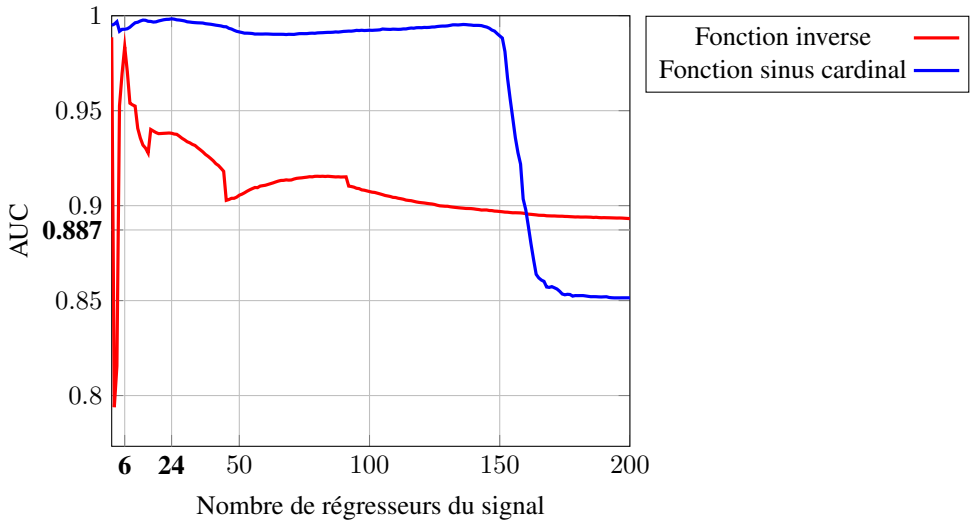


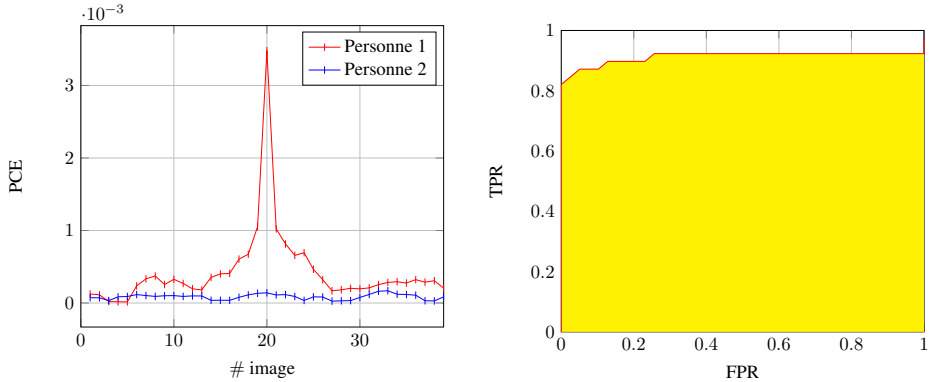
FIGURE 3.22 – Evolution de l'aire sous la courbe (AUC) en fonction du nombre de régresseurs utilisés pour la modélisation du signal pour la fonction sinus cardinal (courbe bleue) et inverse (courbe rouge) tridimensionnelles.

Nous pouvons observer, dans un premier temps, que les deux méthodes permettent une amélioration significative de la classification par rapport à une corrélation sans débruitage. En effet, mis à part le débruitage avec une série de 2 fonctions inverse et avec plus de 160 signaux avec la fonction sinus cardinal, l'aire sous la courbe est supérieure à celle obtenue par notre modèle linéaire pour les deux fonctions de modélisation. Dans un second temps, nous observons une meilleure classification avec la fonction inverse tridimensionnelle, ce qui corrobore notre conclusion de la partie précédente. La fonction inverse présente une AUC maximum de $0{,}980$ pour une série de 6 signaux modélisés, une valeur qui est inférieure à celles présentées par la fonction sinus cardinal pour des séries de 1 à 48 signaux. En outre, la fonction sinus cardinal permet d'obtenir des résultats beaucoup plus stables que la fonction inverse, avec une AUC moyenne de $0{,}985$ pour des séries de 1 à 150 signaux (écart-type de $0{,}002$). La fonction inverse, quant-à-elle, présente une AUC qui diminue asymptotiquement à partir de la série de 6 images pour tendre vers le résultat obtenu sans débruitage du plan de corrélation. De plus, nous observons une diminution brusque de l'aire sous la courbe pour la fonction sinus cardinal entre les séries de 150 et 175 signaux de modélisation pour descendre à une AUC de $0{,}990$ pour une série de 200 signaux, soit une valeur inférieure à la corrélation sans débruitage. Ce comportement s'explique par la taille des signaux modélisés qui, à partir de 150, présentent un lobe central atteignant les dimensions du plan de corrélation. Finalement, nous observons un maximum pour la courbe représentant la fonction sinus cardinal tridimensionnelle pour une série de 24 signaux modélisés, engendrant une $AUC = 0.998$. Nous modéliserons donc notre signal à l'aide d'une série de 24 sinus cardinaux tridimensionnels par la suite.

3.3 Evaluation du débruitage

Pour expérimenter notre algorithme, nous commençons par appliquer notre modèle sur la corrélation des personnes 1 et 2 de la base PHPID. Les séries comportent 39 images. Le filtre VLC utilisé, un filtre de phase pure a été créé à partir de l'image 20 de la personne 1 (visage de face). Les plans de corrélation ont été centrés autour du pic de corrélation suivant la méthode explicitée précédemment. Les régresseurs de bruits comprennent 5 plans de corrélation pour chacune des trois premières personnes de la base PHPID, choisis individuellement pour représenter le plus exhaustivement la diversité des plans de corrélation. Les pics de corrélation ont été supprimés en mettant à zéro une région centrée sur le plan de corrélation de $70 \times 70px$ pour chacun des bruits issus d'une vraie détection (corrélation avec les images de la personne 1). Le signal quant-à-lui a été modélisé à l'aide de la fonction sinus cardinal tridimensionnelle. Une série de 24 signaux a été créée.

Nous présentons les courbes PCE et ROC pour une corrélation sans débruitage du plan de corrélation (les plans originaux) en figure 3.23 et les plans traités avec notre méthode (Fig. 3.24). Nous pouvons tout d'abord remarquer sur les courbes PCE que la méthode de débruitage par décomposition en modèle linéaire permet un rehaussement des niveaux de PCE (Fig. 3.23a et 3.24a) pour les orientations éloignée de l'image de référence (images 1 à 8). De plus, les valeurs de PCE pour la corrélation fausse sont également réduites (notable pour les images 31 à 35). Cet effet engendre une meilleure discrimination entre les deux classes. En effet, on retrouve ce résultat sur les courbes ROC (Fig. 3.23b et 3.24b) : on observe un taux de vraie détection de $74{,}4\%$ avant et $82{,}1\%$ après débruitage pour un taux 0% de fausse alarme sur la courbe ROC classique. Notre méthode de débruitage permet d'obtenir un taux de 100% de vraie détection pour $5{,}1\%$de fausses alarmes, contre 100% en l'absence de débruitage, ce qui dénote une meilleure capacité de notre méthode à différencier les plans de corrélation vraie des plans de corrélation fausse. De plus, l'aire sous la courbe est sensiblement supérieure avec notre méthode de débruitage, passant de $0{,}913$ avant débruitage à $0{,}993$ après débruitage.



(a) (b)

FIGURE 3.23 – Résultats de la corrélation de la personne 1 (image 20) avec les personnes 1 et 2 de la base PHPID : (a) courbes PCE ; (b) courbe ROC.

Une étude de notre algorithme sur l'ensemble des 15 personnes de la base PHPID a ensuite été effectuée, l'image 20 de la personne 1 est utilisée en image de référence d'un filtre de phase pure. La construction du modèle linéaire est inchangée.

La figure 3.25 contient les courbes ROC 3.25a avant et après débruitage 3.25b. On observe, pour un taux de fausse alarme à 0% un taux de reconnaissance plus élevé sans utilisation de notre méthode de débruitage. En effet, celui-ci est de $0{,}821$ sans débruitage tandis qu'il est de $0{,}744$ avec débruitage. En contrepartie, le taux de reconnaissance est de 100% pour $1{,}15\%$ de fausse alarme lors de l'utilisation du débruitage, celui-ci

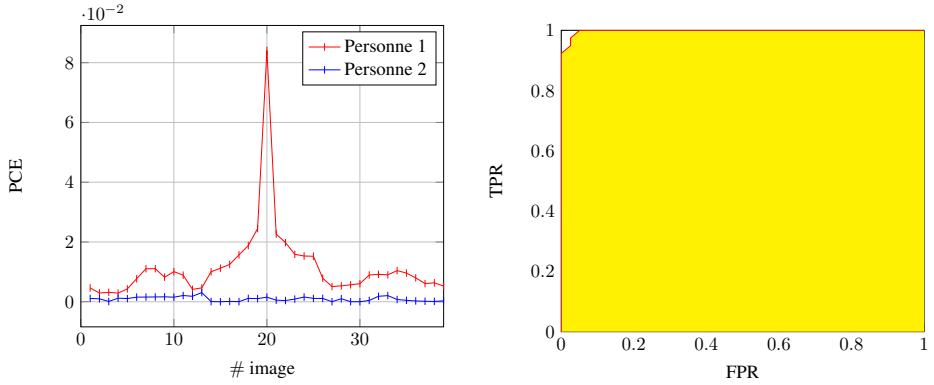


FIGURE 3.24 – Résultats de la corrélation de la personne 1 (image 20) avec les personnes 1 et 2 de la base PHPID, après débruitage du plan de corrélation avec notre méthode : (a) courbes PCE ; (b) courbe ROC.

étant de $72{,}3\%$ avec la corrélation simple. Cet apport apparaît clairement avec l'observation de l'aire sous la courbe, passant de $0{,}887$ sans débruitage à $0{,}998$ avec application de notre modèle linéaire.

Notre méthode permet une reconnaissance du sujet de 100% pour seulement $1{,}15\%$ de fausse alarme et est donc à même d'accentuer la discrimination entre deux classes distinctes et donc le processus de prise de décision avec la base PHPID.

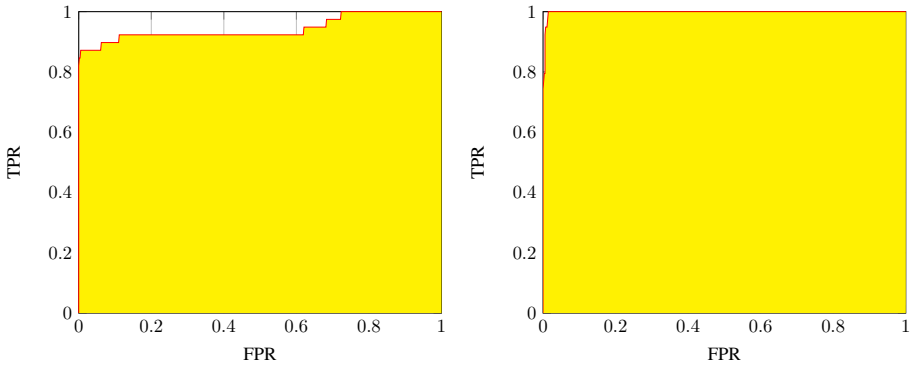


FIGURE 3.25 – Courbes ROC pour une corrélation sur l'ensemble de la base PHPID : (a) corrélation seule ; (b) corrélation avec débruitage.

3.4 Conclusion

Dans ce chapitre, nous avons présenté une méthode de décomposition du plan de corrélation en modèle linéaire. Pour ce faire, nous avons tout d'abord décrit la fabrication des différents régresseurs du bruit et du signal, avec la définition de deux fonction de modélisation du pic de corrélation : la fonction sinus cardinal et la fonction inverse tridimensionnelles. Enfin, nous avons appliqué cette méthode à une étape de débruitage

du plan de corrélation en le reconstituant à l'aide du modèle linéaire en faisant l'omission des régresseurs du bruit, dans le but d'améliorer les performances d'identification de la corrélation VLC.

Dans un second temps, nous avons vu l'importance d'obtenir un pic de corrélation centré dans le plan pour une utilisation du modèle linéaire, les régresseurs composant le modèle du signal étant eux-mêmes centrés. Une étape de découpage du plan autour du pic de corrélation a été définie, permettant un centrage de ce dernier.

Une comparaison des fonctions de modélisation du pic de corrélation a été effectuée. Celle-ci nous a démontré la meilleure capacité d'extraction du signal par la fonction sinus cardinal tridimensionnelle. Le choix de cette fonction de modélisation a enfin permis la création d'un modèle linéaire plus proche de notre signal et donc plus pertinente et plus à même de réaliser un débruitage efficace du plan de corrélation.

L'évaluation de l'effet de la population des régresseurs du signal a enfin permis d'écarter définitivement la fonction inverse tridimensionnelle et de déterminer la taille de série des modèles du signal engendrant une classification optimale.

Nous avons ensuite procédé à une évaluation des capacités de débruitage de notre modèle linéaire. Celle-ci a été accomplie en appliquant un filtre POF et la décomposition en modèle linéaire sur deux personnes de la base PHPID ainsi que sur sa totalité. La comparaison des résultats nous a permis d'apprécier la capacité du débruitage à améliorer la discrimination entre les classes et donc l'étape de prise de décision. Cette étape a éclairé l'apport significatif de la méthode de décomposition en modèle linéaire pour le débruitage du plan de corrélation. Néanmoins, le choix des régresseurs du bruit est une étape délicate influant grandement sur les performances de décomposition du modèle. La création du modèle du bruit nécessite une sélection manuelle des plans de corrélation pertinent. La réalisation d'une méthode automatique permettrait d'augmenter grandement la souplesse de notre algorithme. Cependant, cette étape intervient uniquement lors de la mise en place du modèle.

Notre approche de décomposition en modèle linéaire pour le débruitage du plan de corrélation a fait l'objet d'une publication dans Optics Letters [85].

Chapitre 4

Application de la corrélation pour le suivi

Sommaire

Après avoir considéré différents systèmes de suivi, nous nous dirigeons vers une approche permettant de prendre en compte des modifications de l'objet suivi au cours du temps (rotations, déformations, changements d'échelle). Le modèle utilisé pour détecter l'objet à suivre doit donc être dynamique – ou évolutif, c'est-à-dire qu'il doit être mis constamment à jour. Notre méthode se base sur un suivi itératif, basé sur une comparaison entre une image de référence (contenant l'objet à détecter) et l'image courante et une mise-à-jour à chaque laps de temps de cette image de référence à l'aide des données extraites de l'image cible. L'algorithme, schématisé en figure 4.1, se déroule de la façon suivante : (i) les données extraites de l'image de référence sont recherchées dans l'image courante I_i ; (ii) la mise en correspondance donne la position de la région dans l'image courante ; (iii) cette région est utilisée pour remettre à jour les données de référence ; (iv) on effectue le même traitement dans l'image suivante. Une telle approche est robuste aux déformations (le système d'acquisition devant être adapté aux caractéristiques du mouvement de l'objet). Cependant, la mise-à-jour systématique des données de référence en utilisant l'objet détecté dans l'image courante entraîne l'inconvénient qu'une erreur survenue à l'étape de la mise en correspondance sera perpétuée dans les images suivantes. Ce problème peut être évité en vérifiant la pertinence de notre région détectée, par exemple à l'aide d'une comparaison d'histogrammes (Partie 4.4.2).

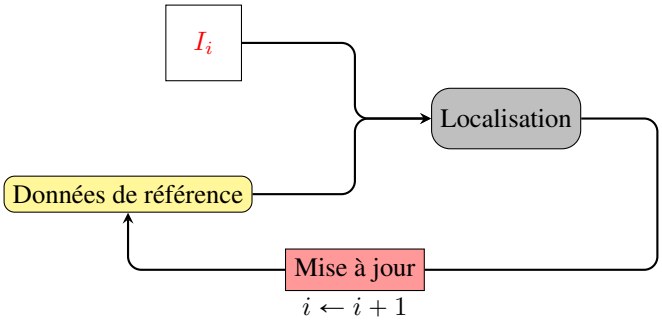


FIGURE 4.1 – Synoptique d'une méthode de suivi itératif.

Comme il a été explicité au chapitre 2, la corrélation se prête particulièrement à une application de suivi itératif. Afin de prendre en compte ses avantages caractéristiques – sa capacité d'effectuer simultanément la détection et le suivi – et d'observer ses performances nous proposons l'utilisation de la méthode JTC (Joint Transform Correlator)[117].

Dans ce chapitre, nous commençons par expliquer le principe de la localisation d'une région dans une image à l'aide du JTC. Nous présentons ensuite son application au suivi itératif. Puis nous décrivons et étudions le rôle des paramètres afin de pouvoir les adapter à notre application : réduction du temps de calcul, optimisation de la précision de localisation. Enfin, nous introduisons une méthode permettant de réinitialiser l'algorithme (comparaison d'histogrammes) et d'optimiser la zone de recherche.

4.1 Localisation à l'aide du JTC

Le principe de la corrélation consiste en une comparaison d'une image cible avec une image de référence (image à reconnaître). Nous nous proposons d'utiliser et d'optimiser un algorithme de suivi basé sur l'architecture JTC (Joint Transform Correlator) [117], caractérisé par la présence combinée de l'image cible et de l'image de référence sur un même plan d'entrée. Comme explicité précédemment au chapitre 2, la méthode se décompose en trois étapes principales :

1. Transformée de Fourier (TF) du plan d'entrée, engendrant le plan de Fourier ;

2. calcul du module au carré du plan de Fourier ;
3. la transformée de Fourier inverse (TF^{-1}) du point 2, permettant d'obtenir le plan de corrélation.

Le plan de corrélation présente trois pics principaux : le pic d'auto-corrélation au centre (ordre zéro) et deux pics périphériques d'inter-corrélation. Ce sont ces deux derniers pics qui nous intéressent ici, parce qu'ils correspondent à la corrélation de l'image de référence avec l'image cible. La position de ces pics dans le plan de corrélation est conditionnée par la position relative de l'image de référence et sa correspondance dans l'image cible.

La figure 4.2 décrit la méthode de localisation de la région recherchée dans l'image cible. L'image de référence et l'image cible sont placées dans un plan d'entrée (Fig. 4.2a), sur lequel est appliqué l'algorithme JTC, renvoyant un plan de corrélation (Fig. 4.2b). La figure 4.2c illustre la localisation des pics sur le plan de corrélation. Les points $A(x_A,y_A)$ et $B(x_B,y_B)$ représentent les deux pics d'inter-corrélation, aux positions (x_A,y_A) et (x_B,y_B) (en pixels) sur le plan de corrélation, respectivement. H correspond à la distance en pixels entre ces deux derniers pics et α à l'angle formé par l'intersection de la droite passant par A et B et de la médiane horizontale du plan de corrélation. Il existe une relation entre ces valeurs dans le plan de corrélation et la position relative de l'image de référence et sa correspondance dans l'image cible (la région de l'image cible corrélant avec l'image de référence) dans le plan d'entrée. On appelle $P_0(x_{P_0},y_{P_0})$ le centre de l'image de référence et $P_1(x_{P_1},y_{P_1})$ son correspondant dans l'image cible. La distance entre P_0 et P_1 est égale à la moitié de la distance entre A et B (i.e. $h = \frac{H}{2}$) – P_0 et P_1 étant équidistants du centre du plan – et l'angle formé par l'intersection de la droite passant par P_0 et P_1 et de la médiane horizontale du plan est égale à α (Fig. 4.2d). Nous obtenons finalement la position de l'objet de référence, recherché dans l'image cible (Fig. 4.2e).

Connaissant la position de l'image de référence P_0 et la localisation P_1 relativement à P_O de sa correspondance, il est possible de retrouver sa correspondante $P(x,y)$ dans l'image cible sur le plan d'entrée, donnée par $x = x_O - x_{P_0}$, $y = y_O - y_{P_0}$, avec $O(x_O,y_O)$ le barycentre du plan d'entrée. Les valeurs de H et de α sont données par l'équation suivante (Eq. 4.1) :

$$\begin{cases} H = \sqrt{(x_B - x_A)^2 + (y_B + y_A)^2} \\ \alpha = \arctan(\frac{y_B - y_O}{x_B - x_O}) \end{cases} \tag{4.1}$$

Nous retrouvons donc la position de la région corrélant avec l'image de référence dans l'image cible (Eq. 4.2) :

$$\begin{cases} x_{P_1} = h\cos(\alpha) + x_{p_0} \\ y_{P_1} = h\sin(\alpha) + y_{P_0} \end{cases} \tag{4.2}$$

Malheureusement, comme énoncé précédemment au chapitre 2, l'architecture JTC souffre de trois inconvénients majeurs : l'ordre zéro (pic d'auto-corrélation), qui est beaucoup plus intense comparativement aux pics d'inter-corrélation ; la faible intensité des pics d'inter-corrélation, rendant le corrélateur sensible au bruit ; et l'évasement des pics d'inter-corrélation, rendant leur localisation approximative avec le JTC classique. La figure 4.3 illustre les plans de corrélation obtenus avec le JTC classique (Fig. 4.3a), le JTC non-linéaire (Fig. 4.3b) et le JTC sans ordre zéro non-linéaire (Fig. 4.3c). L'emplacement théorique des pics de corrélation est mis en évidence par la marque rouge sur les plans. On observe tout d'abord une complète dissimulation des plans d'inter-corrélation sur le plan de corrélation retourné par le JTC classique (Fig. 4.3a), ceux-ci étant très peu puissants comparativement au pic d'auto-corrélation.

Afin de résoudre les problème de la faible intensité et de l'évasement des pics d'inter-corrélation, Javidi [119] propose d'introduire une non-linéarité dans le plan de Fourier par l'intermédiaire d'une élévation à la puissance du spectre joint à l'aide d'un coefficient dit de "non linéarité" $k < 1$. La valeur de ce coefficient k influence fortement la performance JTC. En effet, un coefficient proche de 1 engendrera des pics d'inter-corrélation larges et de faible intensité, tandis que les pics seront nets et intenses avec un k proche de 0. Cela aura donc une incidence sur le comportement du JTC : de grands pics d'inter-corrélation augmentent

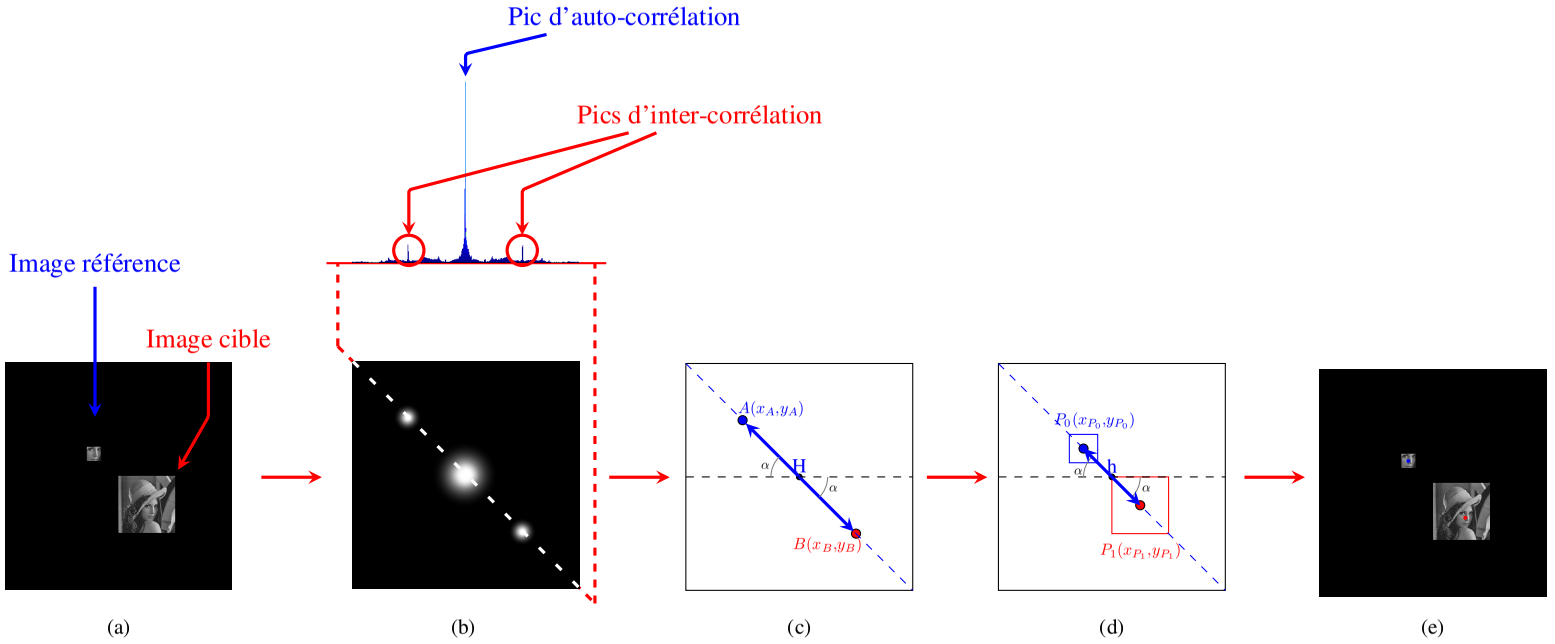


FIGURE 4.2 – Méthode de localisation JTC : (a) plan d'entrée (le point bleu représente le centre de l'image de référence) ; (b) plan de corrélation (en deux dimensions et en coupe) ; (c) localisation des pics de corrélation ($A(x_A,y_A)$ et $B(x_B,y_B)$ représentent les positions des pics d'inter-corrélation, H la distance entre les pics d'inter-corrélation et α l'angle formé par la droite passant par A et B et la médiane horizontale du plan) ; (d) principe de localisation de la correspondance de l'image de référence dans l'image cible ($P_0(x_{P_0},y_{P_0})$ et $P_1(x_{P_1},y_{P_1})$ représentent la position du centre de l'image de référénce et de son correspondant dans l'image cible, h la distance entre les points P_0 et P_1 et α l'angle formé par la droite passant par P_0 et P_1 et la médiane horizontale du plan) ; (e) localisation sur le plan d'entrée (le point bleu représente le centre de l'image de référence et le point rouge la position de son correspondant dans l'image cible).

la robustesse JTC au détriment de sa discrimination et la précision de sa localisation. L'application d'une non-linéarité dans le spectre joint est illustrée en figure 4.3b. Contrairement au JTC classique, les pics d'inter-corrélation sont ici discernables. Malheureusement, il sont encore relativement peu puissants comparativement à l'ordre zéro.

Un compromis entre la robustesse d'une part et la discrimination et l'emplacement précis de l'autre côté doit donc être trouvé. Une étude de l'effet du coefficient de non-linéarité est réalisée en partie 4.3.3.

Il est ainsi nécessaire d'annuler l'ordre zéro. Pour cela nous devons soustraire les auto corrélations du plan de corrélation. Cela est effectué en créant deux plans d'entrée supplémentaires, présentant séparément les cibles et référence des images. Nous déduisons ensuite indépendamment leur transformée de Fourier du spectre joint (Fig. 4.3c).

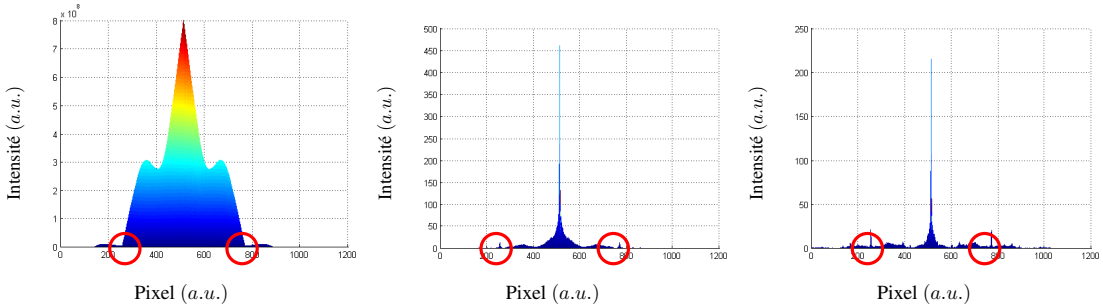
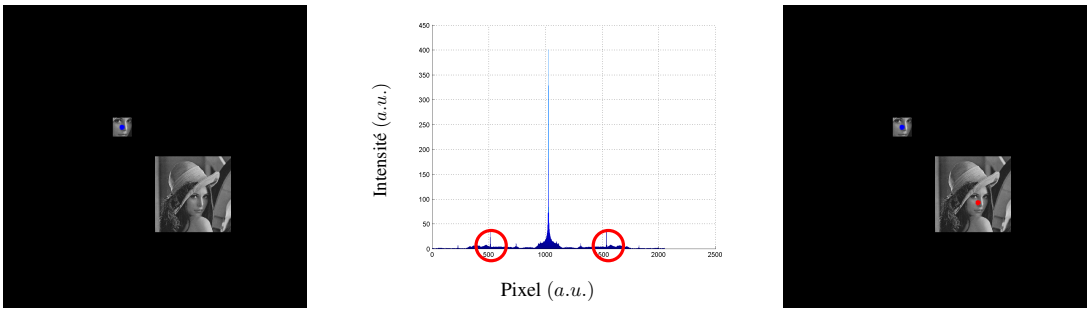


FIGURE 4.3 – Exemples de plans de corrélation : (a) JTC classique ; (b) JTC non-linéaire ($k = 0{,}4$) ; (c) JTC non-linéaire sans ordre zéro ($k = 0{,}4$).

Nous illustrons en figures 4.4 et 4.5 la méthode de localisation JTC. La figure 4.4 présente le cas où l'image de référence et l'image cible sont toutes deux issues de la même image d'origine. Nous commençons donc par créer le plan d'entrée (Fig. 4.4a) : l'image de référence est positionnée sur le quart de plan en haut à gauche du plan, l'image cible sur le quart de plan en bas à droite. Comme énoncé précédemment (chapitre 2), une zone doit être conservée entre les bords du plan d'entrée et les images référence et cible pour éviter un recouvrement de spectre [86]. Nous utilisons ensuite l'algorithme JTC sans ordre zéro et avec un coefficient de non-linéarité $k = 0{,}4$. Le plan de corrélation obtenu est présenté en figure 4.4b. Nous pouvons clairement observer ici deux pics d'inter-corrélation, d'une valeur d'intensité de $35{,}9$, aux positions $A(x_A,y_A) = (513{,}513)$ et $B(x_B,y_B) = (1537{,}1537)$, respectivement. On utilise un plan de $2048px$ de côté. En appliquant l'équation 4.1 nous obtenons $\alpha = 45°$ et $H \simeq 1448px$. Connaissant $P_0(x_{P_0},y_{P_0}) = (804{,}826)$ nous sommes donc en mesure de déterminer la position du visage dans l'image cible : $P_1(x_{P_1},y_{P_1}) = (1317{,}1338)$ (Fig. 4.4c).

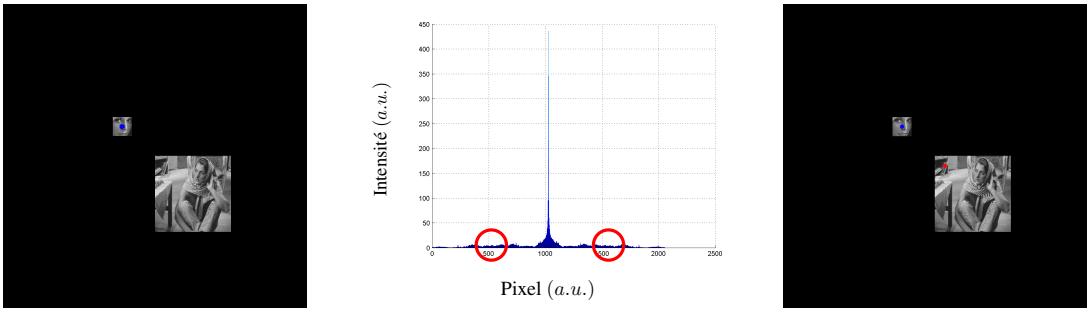
La situation illustrée par la figure 4.5 est celle où l'image cible est différente de celle d'où est issue l'image de référence. Le plan de corrélation issu du plan d'entrée (Fig. 4.5a), observable en figure 4.5b ne comporte aucun pic d'inter-corrélation remarquable. L'image de référence ne corrèle avec aucune partie de l'image cible. Par conséquent les maxima détectés correspondent au bruit engendré par le fond de l'image et la position calculée ne concorde pas avec la position du visage du sujet dans l'image cible (Fig. 4.5a).

À l'aide de l'architecture JTC, nous sommes donc à même de détecter la présence d'une région similaire à l'image de référence dans l'image cible et de la localiser. La corrélation nous renvoie des pics d'inter-corrélation dont le niveau d'intensité nous renseigne sur le degré de similarité entre l'image cible et l'ensemble de l'image. La corrélation étant effectuée sur l'ensemble de l'image cible, en cas de différence notable entre l'image de référence et l'objet à rechercher dans l'image cible, elle peut être plus importante pour une région



(a) (b) (c)

FIGURE 4.4 – Localisation JTC lorsque l'image de référence est inclue dans l'image cible : (a) plan d'entrée (le point bleu représente le centre de l'image de référence) ; (b) plan de corrélation ; (c) localisation de la correspondance de l'image de référence dans l'image cible (le point bleu représente le centre de l'image de référence et le point rouge la position de son correspondant dans l'image cible).



(a) (b) (c)

FIGURE 4.5 – Localisation JTC lorsque l'image de référence n'est pas inclue dans l'image cible : (a) plan d'entrée (le point bleu représente le centre de l'image de référence) ; (b) plan de corrélation ; (c) localisation de la correspondance de l'image de référence dans l'image cible (le point bleu représente le centre de l'image de référence et le point rouge la position de son correspondant dans l'image cible).

correspondant au fond de l'image cible, nous renvoyant une localisation faussée. Deux images de l'objet cible doivent donc être sensiblement similaire pour nous renvoyer une position pertinente de l'objet. La dépendance de la position des pics d'inter-corrélation relativement à la position relative de l'image de référence et de l'objet détecté dans l'image cible est la propriété nous permettant une localisation de cet objet dans l'image cible. Cette localisation de l'objet engendre une capacité d'adaptation de l'algorithme JTC à un suivi itératif d'un objet dans une série d'images.

4.2 Suivi vidéo à l'aide du JTC

4.2.1 Principe

Comme nous venons de le voir, le JTC permet de localiser une image de référence dans une image cible. Il devient dès lors possible de l'utiliser dans un algorithme de suivi, afin de déterminer la trace d'un objet dans une séquence vidéo. Ainsi, nous intégrons l'architecture JTC dans un système de suivi itératif. Comme expliqué précédemment en figure 4.1, page 82, le suivi itératif consiste en une utilisation de la région détectée au temps t comme image de référence au temps $t + 1$. Notre algorithme est illustré en figure 4.6. Au temps t, nous appliquons l'architecture JTC en intégrant dans le plan d'entrée une image référence de l'objet suivi I_{Ref}. Cet objet est recherché dans une image cible I_i, correspondant à la $i_{ème}$ image de la séquence vidéo, intégrée elle aussi dans le plan d'entrée du JTC. Nous obtenons donc en sortie du JTC un plan de corrélation, permettant, comme nous venons de le voir, de localiser une correspondance de l'image référence dans l'image cible. La position ainsi calculée nous permet donc d'extraire la région de l'image cible dans laquelle l'objet de suivi est potentiellement présent, de mêmes dimensions que l'image de référence au temps t. Pour terminer, cette région délimitée dans l'image cible remplace l'image de référence utilisée au temps t pour l'itération suivante (temps $t + 1$), l'image cible étant l'image suivante de la séquence vidéo. Cette architecture auto-adaptative est par conséquent capable de suivre un objet se déplaçant dans l'espace avec des transformations d'échelle ou de point de vue. En effet, l'utilisation d'un matériel d'acquisition adapté aux caractéristiques du mouvement de l'objet qui nous intéresse permet de limiter les déformations de cet objet entre deux images successives. L'image référence est donc mise à jour au fur et à mesure des transformations de l'objet (e.g. échelle).

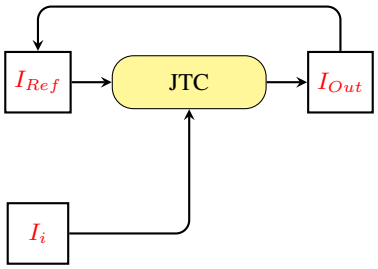


FIGURE 4.6 – Synopsis de l'algorithme itératif par corrélation (Joint Tranform Correlator). I_i correspond à la $i^{ème}$ image de la séquence vidéo, I_{Ref} à l'image de référence et I_{Out} à l'image extraite de I_i.

L'utilisation d'une méthode itérative de suivi basée sur le JTC nécessite la connaissance d'une première image de référence. Il est par conséquent nécessaire de compléter l'algorithme présenté en figure 4.6 par une étape d'initialisation de I_{Ref}. Cette étape peut être effectuée manuellement ou automatiquement. Dans le cas de la méthode manuelle, un opérateur sélectionne une image du visage de la personne à suivre. Cette image est sauvegardée dans le système et est utilisée à chaque initialisation de l'algorithme. Pour notre application, nous utilisons une méthode automatique afin de s'affranchir de l'intervention d'un opérateur, le détecteur proposé par Paul Viola et Michael Jones en 2001 [122] et complété par Rainer Lienhart et Jochen Maydt [123].

La figure 4.7 illustre notre algorithme de suivi itératif JTC complété par l'étape d'initialisation. La partie suivi vidéo est celle présentée en figure 4.6. L'initialisation consiste en une détection et une localisation de l'objet à suivre (ici, le visage d'une personne) à partir de la première image de la séquence (I_0) à l'aide du détecteur de P. Viola et M. Jones. La région détectée est utilisée comme image de référence pour la partie suivi vidéo.

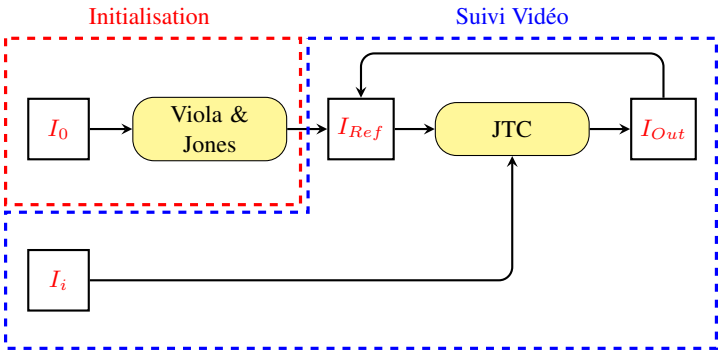


FIGURE 4.7 – Synopsis de l'algorithme itératif par corrélation (Joint Tranform Correlator). I_0 correspond à l'image d'initialisation, I_i à la $i^{ème}$ image de la séquence vidéo, I_{Ref} et I_{Out} à l'image extraite de I_i.

4.2.2 Expérimentation

Un exemple de suivi itératif à l'aide de notre algorithme JTC est illustré en figure 4.8. Les résultats ont été obtenus en utilisant une séquence vidéo de 460 images ($640 \times 480px$, 30 images par seconde). Cette séquence présente une personne positionnée face à la caméra, station debout. Deux déplacements brutaux verticaux sont présents. Le premier se produit de l'image 13 à l'image 29 ($\sim 1s$). Le second déplacement brutal, d'une durée d'environ $200ms$, commence à partir de l'image 106 et se termine à l'image 114.

Afin d'être à même de déterminer la précision de notre algorithme, il nous est nécessaire de disposer d'une réalité terrain du suivi. Pour ce faire, la position de l'objet suivi (dans cette séquence, un visage) a été manuellement déterminée pour chaque image de la séquence par deux opérateurs. La trace de l'objet correspond finalement à la moyenne arithmétique des positions données par les deux opérateurs.

La figure 4.8 présente la distance géométrique en pixels entre la position du visage obtenue avec notre algorithme de suivi et la position annotée manuellement, suivant le numéro de l'image dans la séquence. La taille de l'image de référence est de 78×78 pixels. Les valeurs obtenues restent inférieures à $78px$, le côté de la région du visage, avec un maximum à 58 pixels pour l'image 106 (la seconde chute, la plus brutale). Nous observons principalement deux fluctuations de la distance avec le suivi manuel autour des deux chutes (images 13 et 106). Dans les deux cas, les valeurs diminuent pour les images suivantes. L'algorithme a donc été capable de détecter et de suivre le visage efficacement pendant toute la durée de la séquence, présentant deux changements brutaux du déplacement.

L'architecture JTC est donc en mesure d'effectuer itérativement un suivi d'objet dans une séquence vidéo. De plus, l'utilisation d'une méthode de détection comme le classifieur de P. Viola et M. Jones permet une initialisation automatique de notre méthode. Différents paramètres de notre algorithme sont accessibles, influant sur le temps de calcul et la précision de la localisation. Pour obtenir un suivi robuste, il est nécessaire d'étudier les effets des différents paramètres.

4.3 Les paramètres du suivi

L'architecture JTC et le suivi itératif présentent plusieurs paramètres influençant fortement la précision de la localisation et le temps de calcul. Afin d'optimiser les performances du suivi JTC, nous étudions les effets des différents paramètres disponibles sur la précision et le temps de calcul, à savoir la décimation 4.3.1, la taille du plan de corrélation 4.3.2 et le coefficient de non-linéarité du JTC 4.3.3. Les résultats ont été réalisés au moyen de la séquence vidéo de 460 images présentée en partie 4.2.2 (page 88), contenant deux chutes, l'une

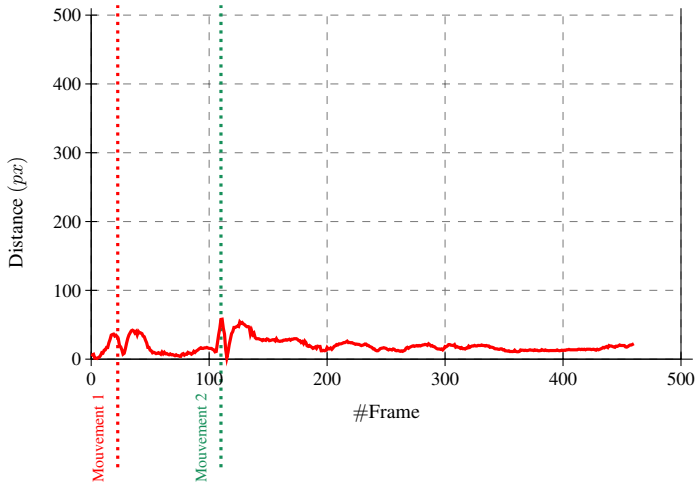


FIGURE 4.8 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images

lente (d'environ $1s$) et l'autre brutale ($\sim 200ms$), débutant aux images 13 et 106, respectivement.

4.3.1 Décimation

La décimation, décrite en figure 4.9 consiste à éliminer une partie des images étudiées. Plus précisément, il s'agit de prendre en compte une plus faible fréquence d'images que celle permise par la caméra utilisée ($30fps$). Une décimation $d = n$, n étant un entier, va conduire à ne prendre en considération que $\frac{1}{n}$ images par rapport à la quantité totale d'images. Ce paramètre a un effet évident sur le temps de calcul et sur les performances de la localisation.

Les figures 4.10 et 4.11 présentent les effets de la décimation sur les performances : la figure 4.10 montre les effets sur la précision du suivi et la figure 4.11 les effets sur le temps de calcul.

La figure 4.10 illustre, pour chaque image de la séquence, la distance relative en pixels entre le suivi JTC et le suivi manuel pour une décimation allant de $d = 1$ à $d = 6$. Lors de l'utilisation de l'ensemble des images de la séquence ($d = 1$), les valeurs de la distance entre le suivi manuel et notre algorithme restent comprises entre $0px$ et $40px$ (l'image de référence, obtenue lors de l'initialisation par le classifieur de P. Viola et M. Jones, est un carré de $78px$ de côté). Le système est donc à même de suivre le visage du sujet durant toute la séquence. Concernant les décimation $d = 2$ et $d > 3$, elles génèrent une perte de la région suivie aux images 106 (chute brutale) et 13 (chute lente), respectivement. Nous observons donc qu'une augmentation de la décimation conduit à une diminution de la précision du suivi. Cela s'explique par un nombre d'images par seconde insuffisant. C'est-à-dire que la différence entre deux images successives devient trop importante pour obtenir des pics d'inter-corrélation énergétiques.

En ce qui concerne les effets de la décimation sur le temps de calcul, les résultats sont présentés sur la figure 4.11, pour une décimation allant de 1 à 6. Il s'agit du temps moyen calculé sur l'ensemble des images de la séquence. Nous observons une augmentation prévisible du temps de calcul lorsqu'on réduit la valeur de la décimation. Celui-ci varie entre $0{,}8s$ pour $d = 1$ à $0{,}14s$ pour $d = 6$.

Comme nous l'avons expliqué précédemment (Fig. 4.10), la décimation affecte grandement la performance du suivi. Un compromis doit donc être choisi entre la performance et le temps de calcul. Ce dernier baisse de façon significative entre les décimations 1 à 3 pour ensuite suivre une tendance asymptotique. Nous nous focaliserons donc par la suite sur une décimation $d \in [1; 3]$.

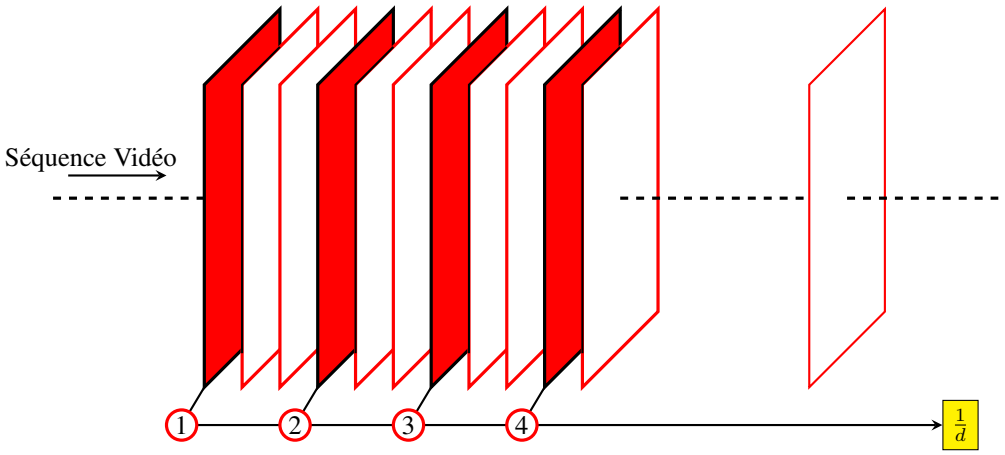


FIGURE 4.9 – Synoptique du principe de la décimation (d correspondant à la valeur de décimation). Les plans représentent les différentes images de la séquence vidéo. Les plans en rouge sont les plans conservés pour le traitement et ceux en blanc, les plans rejetés. Un plan sur trois sont conservés pour cet exemple, correspondant à une décimation $d = 3$.

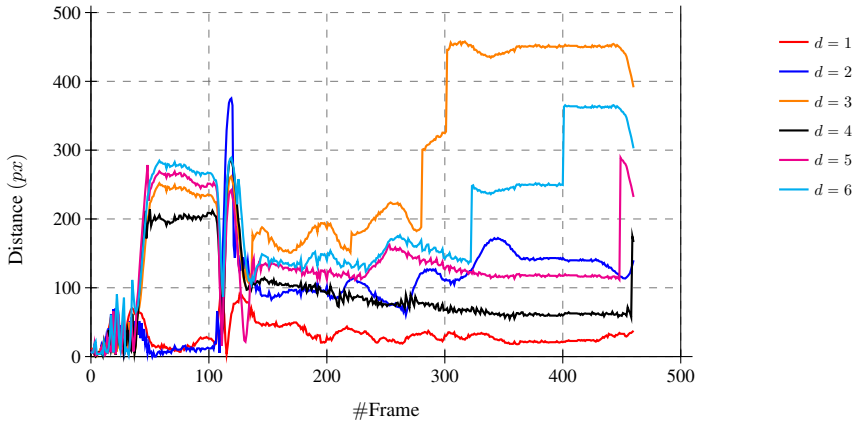


FIGURE 4.10 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour une décimation $d \in [1; 6]$.

4.3.2 Taille du plan de corrélation

La taille du plan de corrélation, dont les effets sont présentés sur les figures 4.12 et 4.13, influe fortement sur la précision de la localisation. De plus, l'architecture JTC non-linéaire sans ordre zéro comporte trois transformées de Fourier et une transformée inverse. La taille du plan engendre donc une répercussion sur le temps de calcul. La figure 4.12 expose la distance en pixels entre le suivi JTC et le suivi manuel pour chaque image de la séquence vidéo, tandis que la figure 4.13 illustre l'effet de cette taille sur le temps de calcul. Les résultats ont été calculés avec un plan de corrélation de dimensions $s_{plane} = (128{,}128)px$, $s_{plane} = (256{,}256)px$, $s_{plane} = (512{,}512)px$ et $s_{plane} = (1024{,}1024)px$

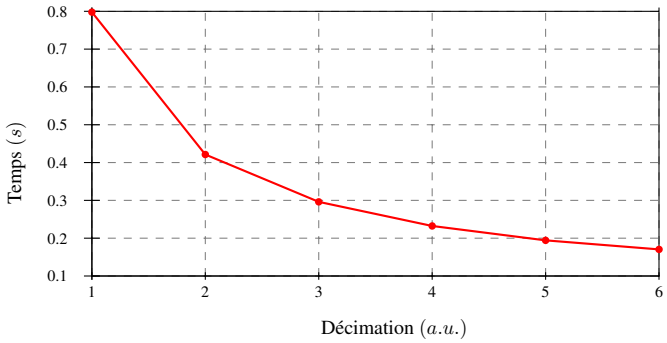


FIGURE 4.11 – Temps de calcul moyen en fonction de la décimation (calculé sur une séquence de 460 images, avec un plan de corrélation de $(512{,}512)px$. C++ / OpenCV, CPU Intel Core ir-2400 3.10 GHz, 4Go RAM, Windows 7 Enterprise 64 bits.

Comme nous pouvons l'observer sur la figure 4.12, les meilleurs résultats de suivi sont obtenus pour une taille de plan de $(512{,}512)px$. Une dimension supérieure entraîne une augmentation du bruit dans le plan de corrélation, tandis qu'une dimension plus faible réduit la précision de la localisation. Par conséquent, cela explique la perte de suivi observée pour les tailles de plan de corrélation de $(128{,}128)px$, $(256{,}256)px$ et $(1024{,}1024)px$.

Finalement, la figure 4.13 présente le temps de calcul par image en fonction de la taille du plan de corrélation. Nous observons que la courbe suit une tendance exponentielle. Cela est dû aux trois transformations de Fourier, nécessaires à l'architecture du JTC non-linéaire sans ordre zéro. Réduire le plan de corrélation est donc essentiel afin d'accélérer le calcul de l'algorithme de suivi. Comme nous le montre la figure 4.12, une taille de plan de corrélation de $(256{,}256)px$ autorise un suivi performant.

4.3.3 Coefficient de non-linéarité

Les résultats présentés sur la figure 4.14 correspondent aux effets du coefficient de non-linéarité (présenté au chapitre 2), générés à l'aide du processus de suivi, pour des coefficients k allant de $0{,}1$ à $0{,}9$. Pour $k = 0{,}3$ et $k < 0{,}5$, nous pouvons observer que les deux courbes varient brutalement aux images 13 et 106, respectivement. La distance par rapport allant d'environ $0px$ à $100px$ environ. La région du visage région faisant $78px$ de côté (dimension obtenue lors de l'initialisation avec le classifieur de P. Viola et M. Jones), le visage du patient est perdu par le procédé de suivi. En effet, aucune de ces deux courbes montrent une diminution significative de la distance après cet événement. Cette perte de suivi est expliquée par la sensibilité du JTC utilisant ces coefficients au bruit, les pics d'inter-corrélation étant pour les deux cas inférieurs aux pics induits par le fond de l'image. En ce qui concerne les valeurs de coefficient $k = 0{,}4$ et $k = 0{,}5$, elles sont similaires, les courbes sur la figure 4.14 étant confondues pendant la majorité de la durée de la séquence. Nous n'observons pas de variation brutale de ces valeurs, ces coefficients permettant un suivi sur l'ensemble de la séquence vidéo expérimentale. Une comparaison avec d'autres séquences vidéo et une observation des images résultantes du suivi nous ont permis de déterminer expérimentalement une valeur optimale du coefficient de non-linéarité $k = 0{,}4$.

4.3.4 Conclusion

Comme explicité précédemment, l'ensemble de ces paramètres affectent fortement le processus de suivi en termes de précision de la localisation et de temps de calcul. Le coefficient de non-linéarité modifie la forme

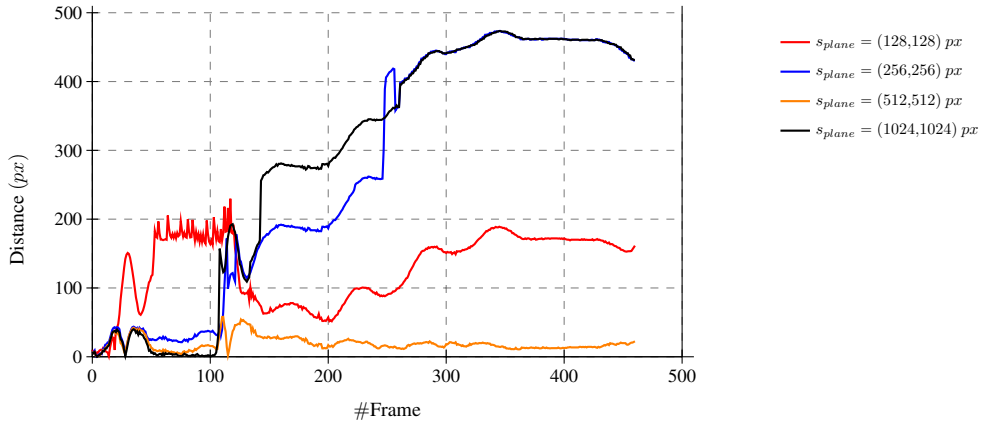


FIGURE 4.12 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour une taille de plan de corrélation $s_{plane} = (2^x,2^x)px$ avec $x \in [7;10]$.

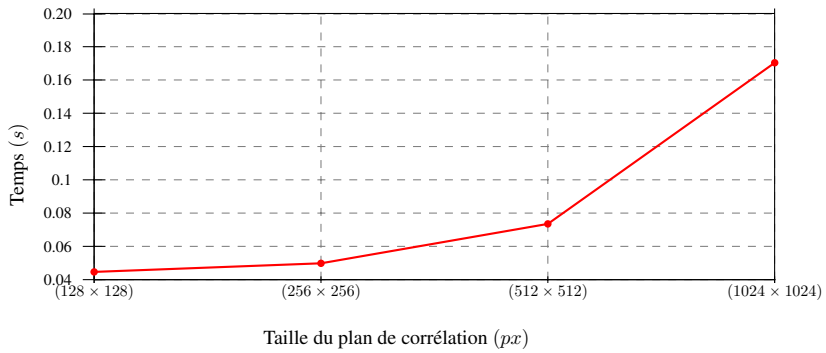


FIGURE 4.13 – Temps de calcul moyen en fonction de la taille du plan de corrélation (calculé sur une séquence de 460 images, avec une décimation $d = 1$. C++ / OpenCV, CPU Intel Core ir-2400 3.10 GHz, 4Go RAM, Windows 7 Enterprise 64 bits.

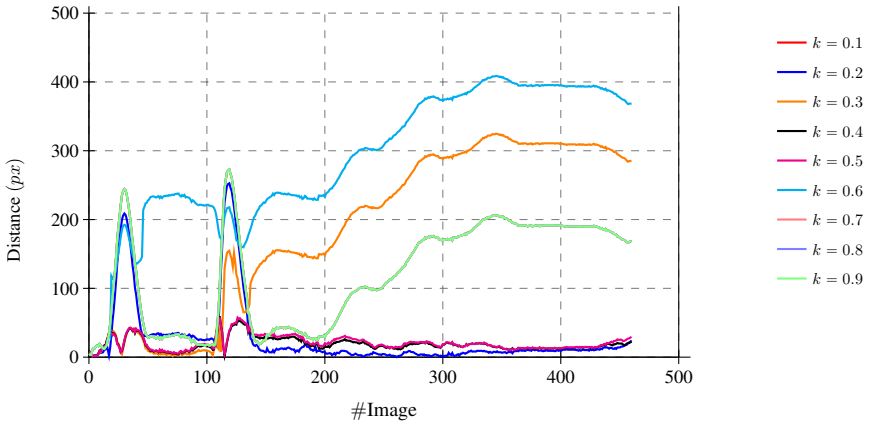


FIGURE 4.14 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour un coefficient de non-linéarité $k \in [0.1; 0.9]$.

des pics d'inter-corrélation, une valeur faible engendrera des pics d'inter-corrélation puissants et localisés et une valeur élevée à des pics évasés. Ainsi, un compromis a été choisi pour k = 0,4. En ce qui concerne la décimation et la taille du plan de corrélation, leurs effets sont de deux ordres agissant à la fois sur la précision de la localisation et de temps de calcul. Par la suite, sauf précision contraire, nous utilisons $k = 0{,}4$ et $s_{plan}(512 \times 512)px$, étant les paramètres donnant les meilleurs résultats de suivi avec le JTC suivi dans nos conditions expérimentales. Ce protocole peut être facilement reproduit dans une autre configuration expérimentale afin d'éliminer une éventuelle déviation de ces paramètres.

4.4 Optimisations du suivi

Malheureusement, le procédé JTC (décrit partie 4.2, page 87) souffre de deux inconvénients majeurs. Tout d'abord, en cas de faible similarité entre la cible et l'image de référence (par exemple, une image floue en raison d'un mouvement rapide), le JTC donne de faibles intensités des pics d'inter-corrélation. Par conséquent, le système devient sensible au bruit. Cette perturbation est d'autant plus susceptible de se produire que nous utilisons une caméra vidéo de faible résolution. Deuxièmement, la localisation dans l'image courante (l'image cible) dépendant uniquement de la bonne localisation de l'image précédente (image de référence), une perte de suivi ne peut pas être détectée ou corrigée par un algorithme itératif. Par conséquent, ce genre d'approches rend définitif toute perte de suivi. Pour remédier à cela, nous proposons deux optimisations : (i) la première est basée sur l'information déterministe : la région à suivre (un visage dans notre cas) ne peut être présent que sur une région spécifique et localisée de l'image cible (i.e. dans le voisinage de sa position précédente), le calcul de corrélation n'est donc utile que sur cette région de l'image ; (ii) la seconde corrélation est effectuée par le calcul d'une mesure de similarité entre l'histogramme de l'image de référence et de la région détectée sur l'image cible. Cette étape supplémentaire permet une détection de la perte du suivi et rend donc possible une ré-initialisation de l'algorithme.

4.4.1 Région d'intérêt

Nous introduisons une première optimisation en réduisant la recherche de l'objet qui nous intéresse (ici, un visage) à une région d'intérêt limitée. L'objectif ici est de contraindre l'algorithme à une région de l'image

pour limiter le bruit induit par le fond de l'image et donc le risque de perte de l'objet suivi. Pour ce faire, nous proposons d'utiliser la connaissance *a priori* de l'emplacement possible du visage. Par conséquent, nous définissons une région d'intérêt dans l'image cible autour de la région de visage détectée dans la trame précédente. Comme représenté sur la 4.15, un facteur échelle est appliqué sur la région du visage $s = (x,y)px$ dans l'image précédente pour obtenir une région d'intérêt, $roi = (a \times s)px$, qui sera finalement utilisée comme image cible pour la corrélation. Travaillant sur une application de détection de chute de la personne âgée, nous sommes surtout intéressés par la partie inférieure de l'image, sous le visage du patient. Ainsi, la région d'intérêt est plus conséquente sur le bas de l'image. Nous conservons tout de même une faible partie au dessus du visage pour des situations inattendues. Nous avons donc adapté ici notre région d'intérêt à un objet spécifique, suivant ses caractéristiques de mouvement. Cette région devra donc être optimisée suivant les caractéristiques de mouvement de l'objet voulu pour chaque application spécifique.

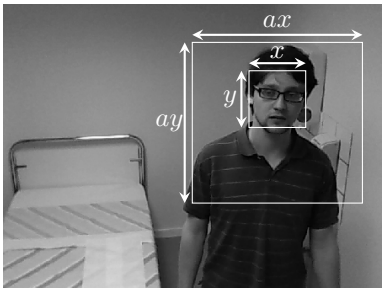


FIGURE 4.15 – Définition de la région d'intérêt ; x et y sont la taille des côtés de la région de visage détectée et a le facteur d'échelle.

On observe sur la figure 4.16 les effets de la région d'intérêt appliquée à image cible. La taille de la région de recherche du visage est définie à l'aide d'un facteur d'échelle appliqué sur la région de visage détectée précédemment (ici, $s = (78,78)px$). Nous observons ici tout d'abord que l'absence de région d'intérêt engendre un suivi très peu fiable. Effectivement, l'objet suivi est perdu dès la première chute (image 13). Lors d'une utilisation d'une région d'intérêt limitée ($s_{roi} = (2 \times s)px$), le mouvement du visage du patient entre deux images peut l'entraîner à sortir de cette région, et donc du champ de vue (mouvements amples et rapides). En ce qui concerne l'utilisation d'une région d'intérêt de grande taille ($s_{roi} = (4 \times s)px$), la réduction du champ peut ne pas être suffisamment élevée pour réduire efficacement l'effet du fond de l'image. En effet, la courbe correspondant présente une variation de distance brusque à l'image 106. Enfin, une région d'intérêt $s_{roi} = (3 \times s)px$ représente le meilleur compromis entre ces deux cas extrêmes, ne conduisant pas à une perte du suivi pour cette séquence.

4.4.2 Correction par similarité d'histogrammes

Comme nous l'avons expliqué précédemment (Partie 4 - 82), le problème principal d'un algorithme itératif réside dans son incapacité à détecter une situation de perte du suivi, c'est-à-dire que l'image de référence corrèle avec, par exemple, une partie du fond de l'image. À l'itération suivante cette partie du fond est donc utilisée comme image de référence. L'algorithme suit donc le fond de l'image et non plus le visage de la personne.

Les pics d'inter-corrélation du JTC ont une intensité dépendante de la ressemblance entre l'image référence et l'image cible. Nous pouvons observer en figure 4.17 les valeurs du PCE en fonction de l'image sur la séquence pour les valeurs de décimation de $d = 1$ à $d = 4$. Nous pouvons observer une chute significative du PCE au moment de la perte du suivi, passant de 0.2496 à 0.238 et de 0.249 à 0.235 à l'image 13 pour les décimations 3 et 4 et de 0.245 à 0.236 à l'image 106 pour la décimation 2. La décimation 1 ne génère aucune

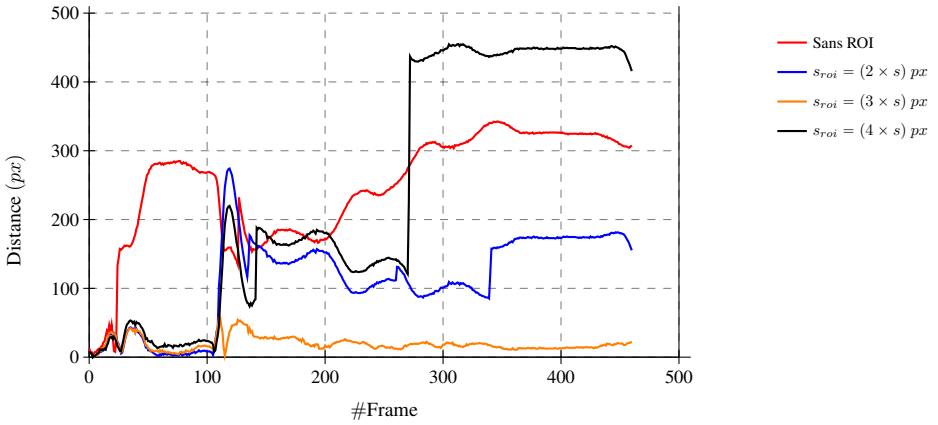


FIGURE 4.16 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour une taille de plan de corrélation $s_{roi} = (a \times s)px$ avec $s \in [2; 4]$.

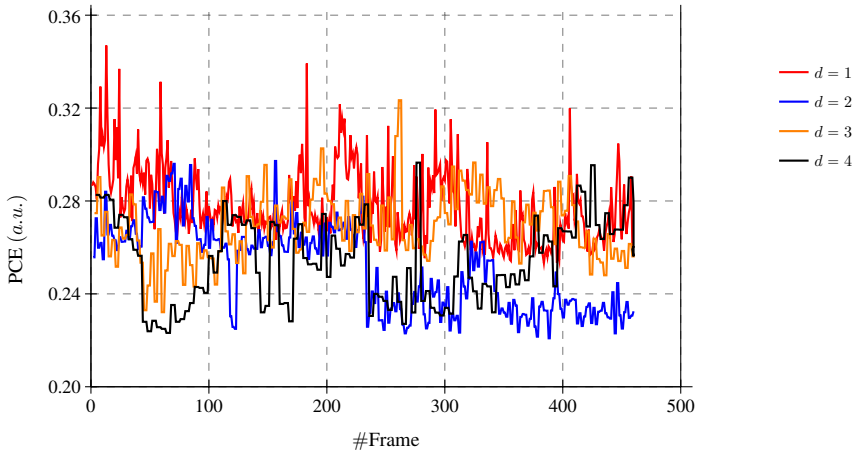


FIGURE 4.17 – Valeur du PCE obtenu avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour une décimation $d \in [1; 4]$.

baisse du PCE lors de ces deux chutes. La vitesse verticale du visage étant beaucoup plus élevée pendant ces événements, il diffère significativement de l'image de référence et ce, d'autant plus que la fréquence de rafraîchissement de l'image de référence est faible (décimation 4). Bien que la valeur du PCE chute lors d'une perte de suivi, nous remarquons également que cette mesure est peu stable, présentant des variations brutales alors même que le suivi continue à se dérouler correctement. La détermination d'un seuil basé sur cette mesure risque donc de rendre le système peu robuste.

Pour remédier à cela, nous proposons de mettre en place une méthode permettant de détecter une telle perte de suivi, en réalisant une comparaison des histogrammes issus de l'image de référence et de la région détectée. Concrètement, dans le cas d'une image, un histogramme est un outil de représentation de la répartition des intensités des pixels. Par conséquent, deux histogrammes provenant de régions similaires visuellement se-

ront relativement proches, tandis que deux régions radicalement différentes nous donneront des histogrammes très dissemblables. En figure 4.18 sont représentés deux histogrammes, l'un provenant du fond de l'image et l'autre de la cible. Nous observons que l'histogramme issu du visage (H_1) contient une distribution relativement étendu (l'ensemble des intensités de 0 à 128). À l'inverse, l'histogramme calculé à partir de la région positionnée sur le fond de l'image (H_2), zone relativement uniforme, nous donne des d'intensités très concentrées au centre.

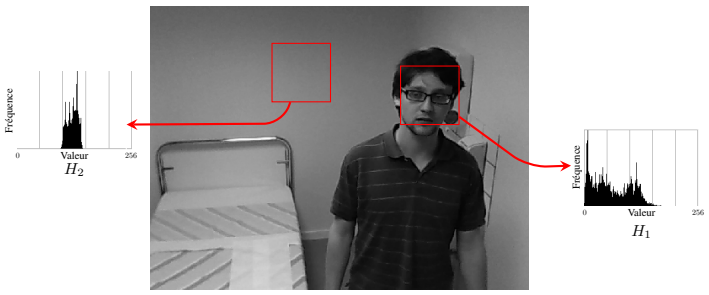


FIGURE 4.18 – Comparaison d'histogrammes.

Afin de différencier deux histogrammes dans des images successives, il est nécessaire de définir un critère de comparaison. Le Chi Square de Pearson [124] est une mesure de non-similarité, définie par l'équation 4.3. Comme expliqué dans l'équation, le Chi Square de Pearson $X^2(H_1,H_2)$ est calculé en additionnant, pour chaque intensité I, la différence au carré de sa densité de probabilité entre la région 1 et la région 2 ($H_1(I)$ et $H_2(I)$, respectivement) divisée par sa densité de probabilité dans la région 1.

$$X^2(H_1,H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)} \tag{4.3}$$

— H_1 et H_2 sont les histogrammes calculés à partir des régions 1 et 2, respectivement ;
— I correspond à la classe de l'histogramme (i.e. intensité).

Dans le cadre d'une application à un système de suivi, on cherche à quantifier la différence entre les histogrammes H_{t-1} et H_t détectée dans au temps t et dans le laps de temps précédent $t-1$ (image de référence). L'équation 4.4 peut donc se réécrire de la façon suivante :

$$X^2(H_{t-1},H_t) = \sum_I \frac{(H_{t-1}(I) - H_t(I))^2}{H_{t-1}(I)} \tag{4.4}$$

— H_{t-1} et H_t sont les histogrammes calcules à partir des régions aux temps $t-1$ et t, respectivement.

Les figures 4.19 et 4.20 illustrent une situation de perte de suivi. Ces résultats ont été obtenus à partir de la séquence vidéo de 460 images décrite précédemment en partie 4.2.2 avec une décimation (voir page 89) $d = 1$ (courbe rouge) et $d = 2$ (courbe bleue), c'est-à-dire $30fps$ et $15fps$, respectivement. La figure 4.19 présente la distance en pixels entre la position du visage obtenue avec notre algorithme de suivi JTC et celle définie manuellement. L'algorithme est donc en mesure de suivre avec précision le visage du sujet pour les deux situations. En ce qui concerne la seconde chute (apparition à l'image 106, durée $\sim 200ms$), nous pouvons clairement observer une brusque variation pour $d = 2$, la courbe allant d'une valeur de distance d'environ $0px$ à une valeur de plus de $200px$, tandis que la ligne rouge, correspondant à $d = 1$, reste à une distance inférieure à $50px$. Cette brusque variation correspond à une situation de perte du suivi. De plus, la distance pour une

décimation de 2 reste élevée suite à cette seconde chute, par comparaison à $d = 1$. Ceci est dû au fait qu'une fois que le suivi est perdu, l'algorithme est incapable de se réinitialiser.

Par conséquent, il est indispensable d'être à même de détecter de telles situations de perte du suivi. La figure 4.20 présente la variation de la mesure de non-similarité des histogrammes successifs, à savoir le Chi Square de Pearson, explicité précédemment. Nous pouvons observer une valeur relativement faible du X^2, aux alentours de 0, dans la situation dite normale (sans décrochage). C'est le cas pour les courbes correspondant à $d = 1$, pour la séquence entière et pour $d = 2$ avant l'image 106 (moment où l'algorithme perd le visage). Pour $d = 2$, nous observons un pic très net (> 6000) à l'image 106. Ainsi, cette mesure de non-similarité est un outil puissant pour la détection de la perte de la région suivie. La courbe bleue revient ensuite à des valeurs faibles, l'algorithme suivant maintenant une partie de l'arrière-plan, les histogrammes des régions détectées sont par conséquent similaires.

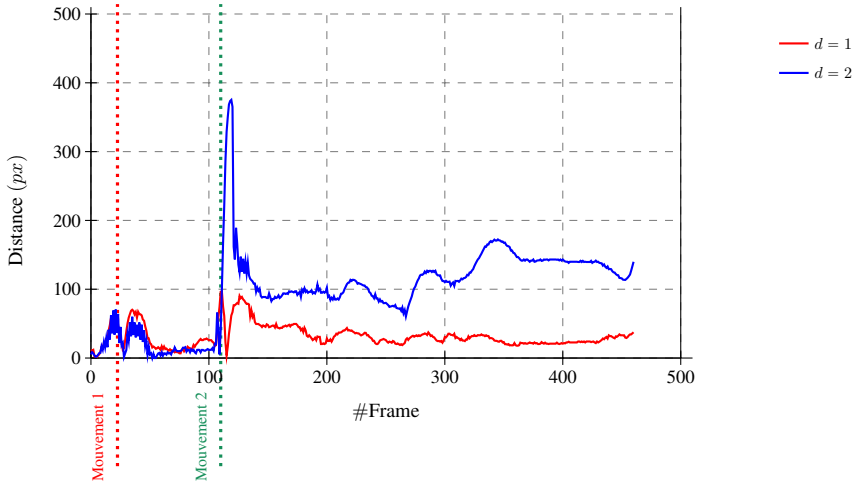


FIGURE 4.19 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images pour une décimation de $d = 1$ (courbe rouge) et $d = 2$ (courbe bleue) avec pour paramètres : $s_{plane} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s = (78{,}78)px$ et $k = 0.4$.

La méthode de comparaison d'histogrammes, explicitée précédemment, est donc à même de détecter une erreur survenue dans le système de suivi JTC. Il devient dès lors possible de réinitialiser notre système lorsqu'une telle situation se présente. L'utilisation de cette méthode dans notre système de suivi est présentée en figure 4.21. À notre algorithme JTC itératif, une étape, appelée mesure de confiance, est rajoutée. À chaque itération, la nouvelle région détectée par le système de suivi JTC est comparée à l'image de référence à l'aide de la mesure de non-similarité d'histogrammes. La valeur du Chi Square de Pearson est comparée à un seuil fixé expérimentalement. Si cette valeur est inférieure au seuil, la région est considérée comme correspondant à l'image de référence. Cette région est donc considérée comme la nouvelle image de référence pour la prochaine itération. À l'inverse si cette valeur est supérieure au seuil, la région détectée ne correspond pas à l'image de référence. Le suivi est donc réinitialisé : on applique le détecteur de P. Viola et M. Jones aux images suivantes de la séquence vidéo jusqu'à détecter à nouveau le visage de la personne, et donc pouvoir le suivre à l'aide de notre algorithme de suivi JTC.

L'application de la correction de notre algorithme est présentée en figure 4.22. Ils ont été obtenus avec une décimation $d = 2$ sur notre séquence de test, avec un seuil $X^2 = 100$. La courbe sans optimisation présente une perte de suivi à l'image 106. La seconde courbe, avec optimisation par la méthode d'histogrammes présente également une augmentation brutale de la distance avec le suivi manuel. Malgré tout, celle-ci reste limitée :

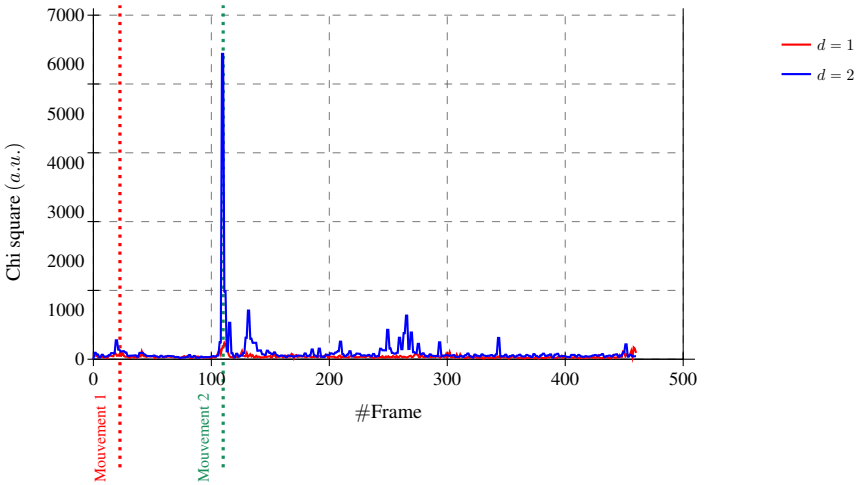


FIGURE 4.20 – Valeur du Chi square de Pearson en fonction du numéro de l'image sur une séquence de 460 images pour une décimation de $d = 1$ (courbe rouge) et $d = 2$ (courbe bleue) avec pour paramètres : $s_{plane} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s = (78{,}78)px$ et $k = 0.4$.

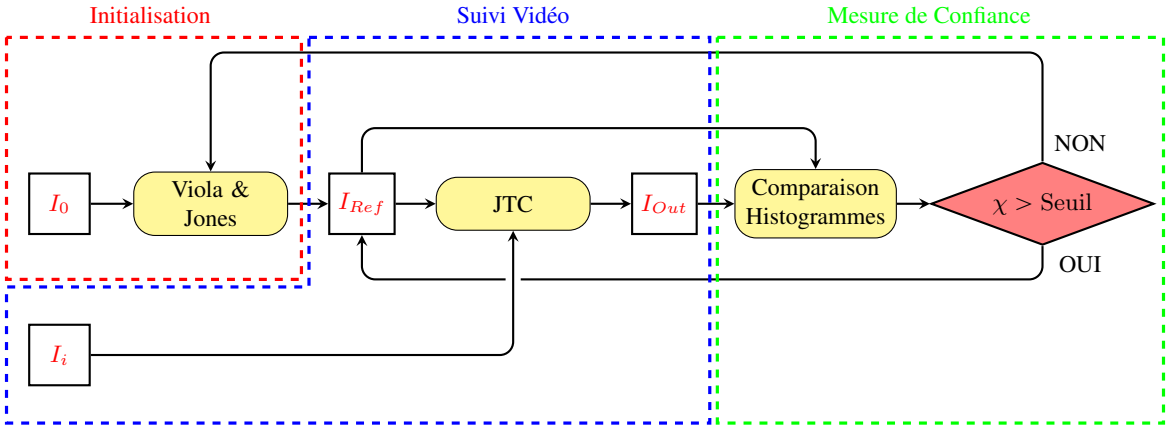


FIGURE 4.21 – Synopsis de l'algorithme itératif par corrélation (Joint Tranform Correlator) avec correction par histogrammes. I_0 correspond à l'image d'initialisation, I_i à la $i^{ème}$ image de la séquence vidéo, I_{Ref} et I_{Out} à l'image extraite de I_i et χ le résultat de notre mesure de similarité d'histogrammes (comparée à un seuil).

$150px$ lors de l'utilisation de la correction alors qu'elle est de $380px$ lorsque l'on n'utilise pas de correction. En effet, le système a perdu l'objet suivi mais les histogrammes de l'image de référence et de la région détectée étaient suffisamment dissemblables pour que notre méthode détecte cette erreur. Le suivi a donc été réinitialisé et l'erreur, corrigée.

Notre optimisation de l'algorithme est à même de détecter une situation de perte de suivi et de la corriger. Cette amélioration permet donc d'obtenir une méthode robuste, capable d'être utilisée pendant un temps relativement long tout en évitant le problème principal des algorithmes itératif, la perpétuation des erreurs générées à chaque itération.

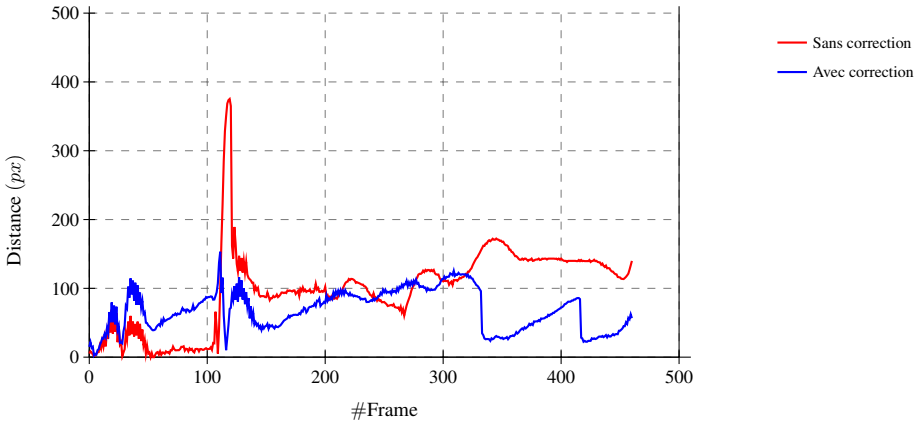


FIGURE 4.22 – Distance en pixels entre la position obtenue par annotation manuelle et celle obtenue avec la méthode de suivi JTC itératif en fonction du numéro de l'image sur une séquence de 460 images avec (courbe rouge) et sans (courbe bleue) utilisation de la méthode de correction par histogrammes, avec pour paramètres : $s_{plane} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s = (78{,}78)px$, $d = 2$ et $k = 0.4$.

4.5 Conclusion

Nous avons présenté dans ce chapitre une application du Joint Transform Correlator (JTC) au suivi itératif d'objet (ici, un visage) dans une séquence d'images. Pour cela nous avons tout d'abord démontré la pertinence du JTC pour la détection et la localisation d'une image de référence dans une image cible. En effet, la position des pics d'inter-corrélation est dépendante de la position relative de l'image de référence et de la région corrélant avec cette dernière. De plus, les améliorations du JTC (suppression de l'ordre zéro et coefficient de non-linéarité) nous permettent de nous affranchir des artefacts engendrés par l'ordre zéro ainsi que d'obtenir des pics puissants et localisés.

Dans un second temps, nous avons proposé une utilisation du JTC pour un suivi itératif et vérifié sa pertinence sur une séquence vidéo. Par la suite, différents paramètres accessibles du suivi (décimation, taille du plan de corrélation, coefficient de linéarité du JTC), affectant les performances ont été étudiés. Pour ce qui est de la décimation et la taille du plan de corrélation, un compromis a dû être choisi entre la précision du suivi et le temps de calcul.

Un protocole d'étude des paramètres de l'algorithme a été mis en place. Le protocole peut facilement être reproductible et les paramètres sont aisément adaptables à l'application de suivi à réaliser (suivi de visage, de cible, de personne...).

Enfin, deux optimisations de notre algorithme ont été apportées. Nous avons tout d'abord défini une région d'intérêt, permettant de limiter la zone de recherche du visage du sujet et donc le risque de corrélation avec une partie du fond de l'image. Ensuite, après avoir observé que les valeurs de la mesure de non-similarité (Chi square de Pearson) entre les histogrammes successifs est grandement affectée par une perte de l'objet suivi, nous avons fusionné cette mesure avec notre suivi JTC itératif pour détecter de tels événements. Cette dernière optimisation est la plus importante car elle nous autorise une détection de cas de perte de l'objet suivi et donc une ré-initialisation de notre algorithme. Par conséquent, elle est à même de contourner l'inconvénient majeur de systèmes de suivi itératifs (cf. page 4), c'est-à-dire leur incapacité à prendre en compte la pertinence de leur données utilisées pour la mise à jour de la référence et donc la perpétuation des inévitables erreurs. Notre algorithme a fait l'objet d'une publication à la conférence Security and Sensing de SPIE [125] et sur SPIE Newsroom [126].

Troisième partie

Application

Chapitre 5

Implantation d'un système de détection de chutes

Sommaire

Le travail de recherche présenté dans ce chapitre expose la mise en place d'une méthode de détection des chutes de la personne dépendante, le problème central d'un système de prise en charge de la personne. Comme nous l'avons explicité au chapitre 1, trois voies majeures se distinguent pour la détection des chutes par vidéo-surveillance : (i) la détection de l'inactivité ; (ii) l'analyse de posture ; (iii) l'analyse des vitesses horizontales et verticales de la tête. C'est cette dernière approche que nous avons choisi, permettant une étude plus poussée de l'activité de la personne.

Nous commençons donc par présenter notre système de détection de chutes de la personne dépendante. Dans un second temps, la méthode d'identification par corrélation présentée au chapitre 3 est expérimentée sur des données adaptées à notre système et comparée avec des méthodes de la littérature. Ensuite, nous développons notre approche de suivi et de détection de chutes. Cette méthode est évaluée à l'aide d'un jeu de données réalisé en conditions réelles et permettant d'explorer les limitations de notre système. Enfin, nous présentons les travaux en cours de finalisation, à savoir la détermination de la profondeur à l'aide de la stéréoscopie.

5.1 Présentation du système

Comme nous l'avons explicité en partie 1.5, notre système se décompose en quatre principaux ensembles (Fig. 5.1) : (i) l'identification ; (ii) le suivi ; (iii) la prise de décision ; (iv) la communication. Par souci de prendre en compte les avantages de la corrélation, que nous avons développée dans le chapitre 2, (l'identification, la détection et le suivi simultanés), nous proposons une méthode de suivi basée sur un corrélateur à spectre joint [117] (Chapitre 4). Enfin, dans le but d'être en mesure d'identifier la personne présente devant la caméra vidéo, afin de se focaliser sur la personne dépendante et d'adapter la réponse en fonction, nous proposons l'utilisation d'un corrélateur de Vander Lugt [95] optimisé (Chapitre 3). Cette optimisation consiste à débruiter le plan de corrélation à l'aide d'un modèle linéaire composé de deux modèles représentant respectivement le signal (le pic de corrélation) et le bruit compris dans le plan. Ce modèle permet finalement de séparer le bruit et le signal afin d'améliorer les performances d'identification.

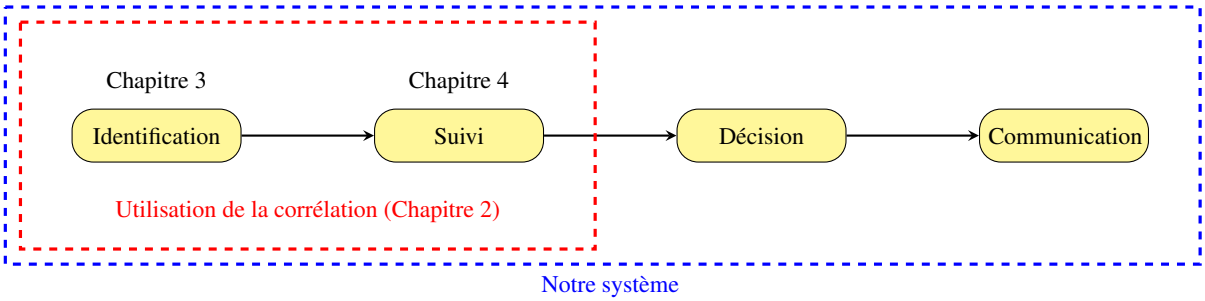


FIGURE 5.1 – Diagramme de notre système de détection des chutes.

Une chute est définie par une perte d'équilibre de la personne, engendrant un affaissement global du corps jusqu'à une position décubitus, caractérisée par la position des jambes, qui ne le soutiennent plus. Cette chute peut survenir brutalement, elle sera alors appelée "chute brutale" ou lentement, une "chute syncopale". La principale étape à caractériser est donc l'action d'affaissement du corps de la personne. Ceci peut être effectué en suivant dans chaque image la position de sa tête, étant le point culminant d'un corps en orthostatisme. En outre, le visage est la partie du corps la plus caractéristique de la personne, contenant les informations les plus riches. Cela est nécessaire pour une focalisation sur la personne d'intérêt lorsque deux individus sont présents sur l'image. De plus, il s'agit de la zone la moins dépendante de l'habillement. Le suivi du visage de la personne nous permet d'obtenir les vitesses horizontale et verticale de cette partie du cours. Les

caractéristiques de la chute constatées par Wu [17] autorisent l'utilisation de ces vitesses pour la mise en place d'un critère de détection de chute. Finalement, la détection d'une chute par notre système entraîne l'émission à l'aide du protocole Simple Mail Transfer Protocol (SMTP) d'un message d'alerte par courrier électronique ou Short Message Service (SMS).

Nous avons dans un premier temps réalisé un système se basant sur l'information en deux dimensions disponible sur l'image. Il s'agit d'une première tentative permettant une optimisation des méthodes d'identification et de suivi. Considérant l'utilisation d'une détection de chute basée sur les vitesses verticale et horizontale de la tête de la personne, il nous est nécessaire de prendre en compte l'information 3D (travaux en cours de réalisation).

Nos conditions expérimentales sont présentées en figure 5.2. Dans un soucis d'expérimentation de notre système dans des conditions d'approchant de l'environnement d'installation réel, nous avons mis en place une chambre de tests (Fig. 5.2a) reproduisant ce que l'on peut trouver notamment dans des centres d'hébergement des personnes dépendantes et âgées. Elle se compose d'un lit, d'un espace de toilette (i.e. lavabo), d'un fauteuil, d'une table basse et d'une télévision, que nous utilisons principalement pour l'affichage de nos données. Dans cette pièce a été disposé notre système détection des chutes, à savoir des caméras vidéo de type webcam (Fig. 5.2b) et une unité centrale. Les spécifications sont résumées sur le tableau 5.1. Les webcams utilisées sont de type Logitech C525 (résolution de $1280 \times 720px$ avec une fréquence de rafraîchissement de $30fps$. L'unité centrale est équipée d'un CPU de $3{,}10GHz$, avec $4Go$ de RAM. Notre système a été implanté pour l'identification en Matlab et pour le suivi en C++, à l'aide de la librairie OpenCV [127].

Nous commençons par présenter la méthode d'identification utilisée. Dans un second temps la méthode de suivi sera développée et appliquée à la détection de chutes.

(a) (b)

FIGURE 5.2 – Conditions expérimentales : (a) chambre d'expérimentation : (b) webcam utilisée (Logitech C525).

5.2 Identification

Notre système de détection des chutes est basé sur le suivi de la tête de la personne. Comme explicité précédemment, l'identification est une étape importante de notre système de détection des chutes, évitant par exemple une initialisation sur une autre personne que l'habitant (e.g. visiteur, enfant). Cela permet également une utilisation en conditions réelles, car rendant possible la présence combinée de plusieurs personnes dans le lieu de vie. Finalement, dans le but d'augmenter les capacités de notre système dans de futurs travaux, cela autorise également l'enregistrement de données personnalisées de l'individu afin d'apporter des précisions

TABLE 5.1 – Spécifications.

Unité centrale	CPU		Intel Cor ir-2400 3,10GHz
	RAM		$4Go$
	OS		Windows 7 Enterprise 64 bits
	Implantation	Identification	Matlab
		Suivi	C++ / OpenCV
Caméra	Modèle		Logitech C525 HD
	Résolution		$1280 \times 720px$
	Focus		Auto
	Rafraîchissement		$30fps$

complémentaires au personnel de santé.

Comme explicité précédemment aux chapitres 2 et 3, la corrélation optique permet une identification efficace d'objets, et est particulièrement adaptée à la reconnaissance de visages. Le corrélateur de Vander Lugt présente l'avantage de nécessiter la définition d'un filtre en amont de l'étape de corrélation, rendant possible une optimisation de ce filtre à la personne et à l'application voulue. En effet, l'utilisation d'un filtre de phase pure (POF) [92] permet d'obtenir une corrélation très discriminante, retournant un taux de fausses alarmes faible au détriment du taux de non-détections fausses. L'approche multicorrélation, quant à elle, est à même de combiner plusieurs images de références dans le même filtre de corrélation. L'objet peut donc être appris selon plusieurs angles de vue ou avec différentes déformations tout en conservant le même temps de calcul de la corrélation que lors de l'utilisation d'un filtre de corrélation classique, permettant d'augmenter le taux de bonne détection tout en maintenant un faible taux de fausses alarmes.

5.2.1 Apport de notre méthode de débruitage

Dans le chapitre 3, nous avons présenté une optimisation des résultats de la corrélation, basée sur un débruitage du plan de corrélation à partir d'une décomposition de ce dernier en modèle linéaire. Cette étape consiste à décomposer le plan de corrélation en une combinaison linéaire de plusieurs composantes prédéfinies, les régresseurs. L'objectif de cette décomposition ici est de séparer le pic de corrélation du bruit présent dans le plan de corrélation afin d'augmenter les performances de la détection. Ainsi, les régresseurs du modèle linéaire sont choisis en fonction : le plan est décomposé selon des régresseurs du bruit et des régresseurs du signal. Les régresseurs du bruit ne pouvant être modélisés virtuellement, ceux-ci sont donc créés à partir de plans de corrélation issus d'une corrélation entre une image référence et cible de deux objets différents et entre une image cible et référence du même objet dont le pic a été artificiellement retiré. Les régresseurs du signal sont, quant à eux, créés à partir d'une fonction sinus cardinal en trois dimensions (une fonction inverse tridimensionnelle a été écartée expérimentalement au chapitre 3).

L'application du modèle ainsi créé sur un plan de corrélation nous autorise à calculer les coefficients permettant une décomposition linéaire de ce plan suivant les régresseurs du bruit et du signal. Une partie du signal, ne pouvant être expliquée par le modèle, est obtenue en faisant la différence entre les plans de corrélation original et reconstruit. Enfin, nous obtenons le plan débruité en additionnant le modèle du signal auquel on applique les coefficients calculés et le bruit résiduel.

Une application de notre méthode sur la base PHPID nous a permis d'apprécier sa pertinence dans un contexte d'identification de personnes (Chapitre 3). Nous présentons dans cette partie une utilisation de cette approche dans le cadre de notre système de détection de chutes de personnes dépendantes.

5.2.1.1 Expérimentation

Afin de représenter les performances de l'identification à l'aide de la corrélation Vander Lugt, il est nécessaire de l'expérimenter en conditions réelles. Comme nous l'avons explicité, notre système de détection des chutes est imaginé pour être implanté dans l'environnement de vie de la personne dépendante. Des images de référence doivent par conséquent y être intégrées lors de l'installation. Ainsi, il nous est impératif de disposer d'une base de données comprenant le visage de l'habitant, ce afin de réaliser les filtres de corrélation, c'est-à-dire pour la phase d'apprentissage du système. Enfin, il est nécessaire que la prise de vue soit effectuée selon divers angles de rotation du visage.

Une base de données de visages a donc été réalisée. Elle est basée sur un principe similaire à celui utilisé pour la base PHPID [116]. Notre base comporte 4 personnes. Les orientations du visage vont de $+45^\circ$ à -45° avec un pas de 15° dans la direction horizontale, et sont de -10°, 0° $+10^\circ$ dans la direction verticale. Une série complémentaire a été enregistrée pour la personne 1, ce afin d'utiliser des images différentes pour les étapes d'apprentissage et de classification.

La corrélation a été effectuée au moyen de filtres composites à partir des images de la série complémentaire de la personne. Six filtres, présentés en figure 5.3 ont été réalisés afin d'éviter les problèmes de saturation inhérents au filtre composite. Ceux-ci comprennent l'ensemble des 21 images de la série (filtre 1 : images 3, 4 et 5 ; filtre 2 : images 1, 2, 6 et 7 ; filtre 3 : images 10, 11 et 12 ; filtre 4 : images 8, 9, 13 et 14 ; filtre 5 : images 14, 17, 18 et 19 ; filtre 6 : images 15, 16, 20 et 21). La fusion des données issues de la corrélation avec les images de la base de tests pour chacun des six filtres a été effectuée en prenant la valeur maximale de PCE obtenue pour chaque image.

Le modèle linéaire, quant à lui, a été construit à l'aide de 24 signaux modélisés avec la fonction sinus cardinal tridimensionnelle. Le modèle du bruit est créé de la même manière que dans le chapitre 3, c'est-à-dire en intégrant au modèle cinq plans de corrélation pour chaque personne de test, choisis de manière à représenter la diversité des bruits de corrélation.

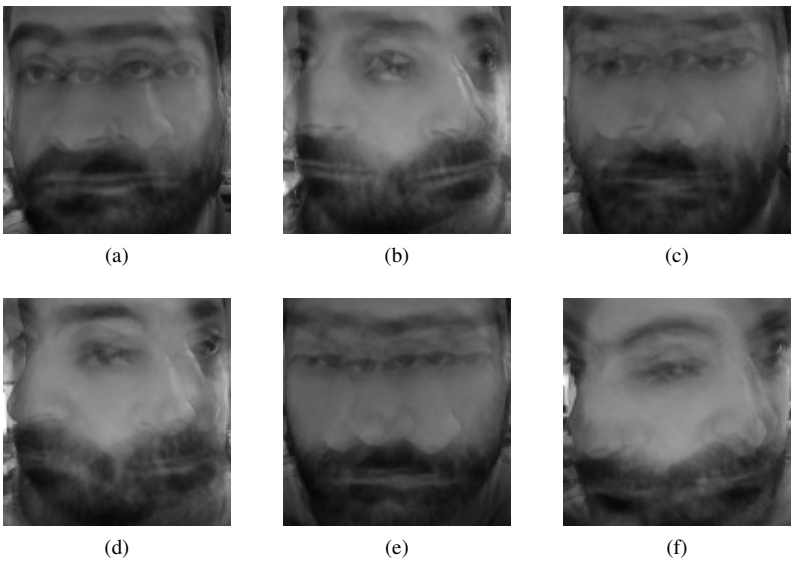


FIGURE 5.3 – Filtres composites utilisés pour la corrélation réalisés avec les images : (a) 3, 4 et 5 ; (b) 1, 2, 6 et 7 ; (c) 10, 11 et 12 ; (d) 8, 9, 13 et 14 ; (e) 14, 17, 18 et 19 ; (f) 15, 16, 20 et 21.

Les courbes ROC pour l'identification par la corrélation de Vander Lugt sont présentées en figure 5.4. Nous

illustrons les résultats de la corrélation sans (Fig. 5.4a) et avec utilisation de notre méthode de débruitage du plan de corrélation à l'aide du modèle linéaire (Fig. 5.4b). Elles ont été générées en corrélant chacun des 6 filtres de la série complémentaire de la personne 1 avec l'ensemble des images de notre base de données. Pour chaque image, la valeur retenue est le PCE maximum obtenu pour les 6 corrélations. La courbe ROC permet de rendre compte de la capacité de classification d'un filtre. Cependant l'une des principales limitations de la courbe ROC tient à leur incapacité à différencier un filtre très discriminant d'un filtre peu discriminant.

Nous observons une amélioration significative de la classification lorsque notre méthode de débruitage est appliquée. En effet, le taux de vrais positifs atteint 100% pour un taux de faux positifs de $3{,}17\%$ avec débruitage. Celui-ci est de $30{,}16\%$ pour la même valeur de TPR pour la corrélation pure. De même, pour un FPR de 0% nous obtenons un TPR de $81{,}00\%$ sans et de $90{,}50\%$ avec débruitage. Ce résultat se retrouve dans l'aire sous la courbe, celle-ci étant de $AUC = 0{,}9811$ pour la corrélation sans débruitage et de $AUC = 0{,}9985$ avec la décomposition en modèle linéaire.

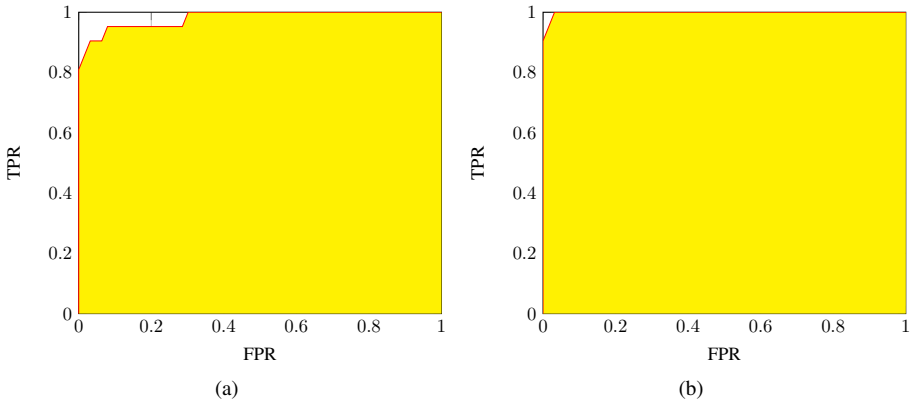


FIGURE 5.4 – Courbes ROC pour une corrélation sur l'ensemble de la base avec les 6 filtres de corrélation (critère : PCE maximum) : (a) corrélation seule ; (b) corrélation avec débruitage.

L'effet du débruitage du plan de corrélation à partir d'une décomposition de celui-ci en modèle linéaire, mis en valeur au chapitre 3, se retrouve donc bien lors de son application sur notre propre base de données. En effet, notre méthode permet bien d'obtenir une classification sensiblement plus performante.

5.2.2 Comparaison de notre méthode avec la littérature

Afin d'évaluer objectivement les performances de l'identification par corrélation de Vander Lugt, et plus particulièrement l'apport de notre méthode de débruitage, il est nécessaire de se situer par rapport à des méthodes classiquement utilisées dans la littérature. Zhao et al. a proposé trois catégorisations des méthodes de reconnaissance faciale [128] : (i) les méthodes basées sur une analyse du visage entier ; (ii) les méthodes utilisant des caractéristiques du visage (e.g. nez, bouche) ; (iii) les méthodes hybrides. La technique de reconnaissance par corrélation se plaçant dans le cadre des méthodes globales, analysant la région du visage dans sa totalité, nous avons choisi de la comparer avec les approches Fisherfaces [129] et Eigenfaces [81] ainsi qu'avec une approche hybride, LBPH (Local Binary Pattern Histograms) [130].

Les méthodes de reconnaissance Fisherfaces et Eigenfaces sont toutes deux basées sur l'analyse en composantes principales (ACP). Le principe de l'ACP, énoncée par K. Pearson en 1901 [131] et H. Hotelling en 1933 [132], consiste à réaliser une transformation de l'espace de représentation des variables d'observation. Il s'agit d'obtenir un ensemble de variables dé-corrélées à partir d'un ensemble plus grand de variables d'entrées

corrélées. Les directions de plus grande variance dans les données d'entrée, donnés par l'application de la méthode PCA sont appelées composantes principales.

L'approche Eigenfaces, proposée par M. Turk et A. Pentland en 1991 [81] est l'application directe de l'analyse en composantes principales à la reconnaissance faciale. Un ensemble de vecteurs orthogonaux est réalisé à partir d'une base d'apprentissage composée d'une série d'images de visages. Le visage à reconnaître est finalement décomposé en une combinaison linaire (somme pondérée) de ces vecteurs. Le plus proche voisin entre l'image d'entrée projetée dans le sous-espace et les vecteurs orthogonaux obtenus à partir de la base d'entraînement permet la prédiction de la personne présente sur l'image d'entrée. La méthode Eigenfaces a l'inconvénient de ne pas regrouper les données d'entrée en classes : chaque image de la même personne est vue comme une donnée indépendante.

À l'inverse, l'approche proposée par R. A. Fisher en 1936 [129], appelée Fisherfaces, et appliquée à la reconnaissance faciale en 1997 par P. Belhumeur et al. [133] est basée sur une analyse linéaire discriminante (ALD). Le principe de l'ALD réside dans une maximisation de la distance entre les matrices inter-classes et sa minimisation entre les matrices intra-classes.

Finalement, la méthode LBPH, proposée par T. Ahonen et al. en 2004 [130], est basée sur la description des caractéristiques locales d'un visage, afin de préserver l'invariabilité de la méthode suivant des changement d'échelle ou de rotation, à partir de la texture de l'image. L'idée est de décrire chaque image de la base d'apprentissage, après l'avoir divisée en régions de taille fixée, à l'aide de la comparaison de chacun des pixels centraux des régions avec ses voisins, ce afin d'extraire des points caractéristiques locaux. Chaque pixel dans l'entourage d'un pixel central est remplacé soit par la valeur 1 s'il est supérieur soit 0 s'il est inférieur. Ces régions binaires ainsi obtenues sont appelées Local Binary Pattern (LBP). Lors de l'étape de reconnaissance, l'image contenant le visage à reconnaître est également analysée de la même façon. Enfin, à partir des LBP est réalisé un histogramme pour chaque image. La plus faible distance entre l'image du visage à reconnaître et les images de la base d'apprentissage nous permet de prédire la personne présente.

Ces trois méthodes ont été choisies pour évaluer l'approche VLC pour la reconnaissance faciale et notre algorithme de débruitage du plan de corrélation à l'aide du modèle linéaire. Pour ce faire, nous avons construit deux bases de données : une base d'apprentissage (Annexe A) et une base d'expérimentation (Annexe B), résumées respectivement sur les tableaux 5.2 et 5.3. La base d'apprentissage, comprend huit personnes différentes. Elle utilise des images de notre base de donnée présentée en partie 5.2.1.1 et des images de la base PHPID [116]. Pour les trois premiers sujets, 21 images par individu ont été prises, suivant des orientations de $-10°$, $0°$ $+10°$ dans la direction verticale et dans la direction horizontale de $+45°$ à $-45°$ avec un pas de $15°$. Les quatre autres personnes sont issues de la base PHPID et comprennent chacune 53 images. Les séries d'images diffèrent des trois précédentes en cela que les poses dans la direction horizontale ont été prises avec un pas de $10°$. Chaque image a été recadrée sur le visage de la personne et redimensionnée à $121 \times 121px$ pour obtenir une cohérence avec les trois premières séries de la base. La base de test, quant à elle, comprend deux sujets différents. Ceux-ci correspondent aux individus 1 et 2 de la base d'apprentissage. Ces deux nouvelles prises de vue ont été réalisée de la même façon que leurs homologues dans la base d'apprentissage, mis-à-part un port de lunettes de vue pour la personne 2 dans la base d'apprentissage.

L'apprentissage a été réalisé pour chacune des méthodes avec la base résumée dans le tableau 5.3. Pour Eigenfaces, Fisherfaces et LBPH, chaque série a été étiquettée en fonction de l'individu présent. Pour la corrélation VLC, 6 filtres ont été réalisés pour les séries de 21 images et 13 pour les séries de 53 images, comme présenté en tableau 5.2. Comme précédemment, le plan de corrélation est recentré autour du pic de corrélation. Le modèle linéaire, quant à lui, a été créé à partir de 24 régresseurs de signal (fonction sinus cardinal tridimensionnelle). Différents modèles de bruit ont été générés :

Bruit ♯1 : pour chaque filtre, un bruit composé de la corrélation du filtre avec l'ensemble de la base d'apprentissage ;

Bruit ♯2 : pour chaque filtre, après corrélation du filtre avec l'ensemble de la base d'apprentissage, seuls les 5 plans de corrélation pour chaque individu de la base donnant le PCE le plus élevé lorsque l'image

TABLE 5.2 – Données d'apprentissage et filtres utilisés en fonction du nombre d'images par sujet, de l'amplitude et du pas de rotation du visage et de la base d'origine (personnelle ou PHPID).

Personne		1	2	3	4	5	6	7	8
Nombre d'images		21	21	21	21	53	53	53	53
Amplitude		45°	45°	45°	45°	45°	45°	45°	45°
Pas		15°	15°	15°	15°	10°	10°	10°	10°
Base d'origine		PERS	PERS	PERS	PERS	PHPID	PHPID	PHPID	PHPID
Filtres	♯1	images 3, 4, 5				images 1, 2			
	♯2	images 1, 2, 6, 7				images 3, 4			
	♯3	images 10, 11, 12				images 5, 6, 7			
	♯4	images 8, 9, 13, 14				images 13, 14, 15			
	♯5	images 17, 18, 19				images 10, 11, 17, 18			
	♯6	images 15, 16, 20, 21				images 12, 16			
	♯7					images 8, 9, 20, 19			
	♯8					images 26, 27, 28			
	♯9					images 23, 24, 30, 31			
	♯10					images 25, 29			
	♯11					images 21, 22, 32, 33			
	♯12					images 39, 40, 41			
	♯13					images 36, 37, 43, 44			
	♯14					images 38, 42			
	♯15					images 34, 35, 45, 46			

TABLE 5.3 – Données de test en fonction du nombre d'images par sujet, de l'amplitude et du pas de rotation du visage et de la base d'origine (personnelle ou PHPID).

Personne	1	2
Nombre d'images	21	21
Amplitude	45°	45°
Pas	15°	15°
Base d'origine	PERS	PERS

cible ne contient pas la personne correspondant au filtre, et les 5 pour lesquels le PCE est le plus faible lorsque l'image cible et le filtre sont issus du même individu sont retenues ;

Bruit ♯3 : pour chaque filtre, seuls 5 plans sont retenus aléatoirement pour chaque individu de la base d'apprentissage (6 tirages sans remise).

Le pic de corrélation est retiré en annulant une région de $30 \times 30px$ au centre du plan de corrélation lorsque l'image cible appartient à la personne utilisée pour la création du filtre.

Le tableau 5.4 contient les résultats de la comparaison des méthodes Eigenfaces, Fisherfaces, LBPH et VLC sans et avec débruitage par application du modèle linéaire (VLC + LM). Les modèles de bruit ♯1 et ♯2 y sont présentés. Pour l'ensemble des méthodes, chaque image test a été comparée avec les données issues de la phase d'apprentissage. À l'issue de cette étape, la personne présente dans l'image a été prédite, nous permettant d'obtenir un taux de bonne reconnaissance et de mauvaise reconnaissance. Pour ce qui est de la méthode VLC et VLC + LM, la personne a été prédite soit en labellisant l'image à partir du filtre avec lequel le PCE est le plus élevé, soit en calculant la moyenne des PCE obtenus en corrélant chacun des filtres d'une

personne avec l'image cible. L'image cible est labellisée en fonction de l'individu de la base d'apprentissage avec lequel sa corrélation offre le score le plus élevé.

Nous pouvons tout d'abord observer un taux élevé de bonne reconnaissance pour les méthodes Fisherfaces et LBPH, de $95{,}24\%$. La reconnaissance avec Eigenfaces est légèrement moins efficace, donnant un taux de bonne reconnaissance de $83{,}34\%$. Cela est dû à une meilleure robustesse aux variations de luminosité et d'expression du visage de la méthode Fisherfaces par rapport à la méthode Eigenfaces, principalement du fait du regroupement des données d'apprentissages en classes [133]. La corrélation VLC quant à elle, est celle offrant la prédiction la plus performante, de $97{,}62\%$ avec l'utilisation des la moyenne des PCE pour chacun des filtres d'une personne de la base d'apprentissage. À l'inverse, notre méthode de débruitage du plan de corrélation semble être peu performante dans cette expérimentation, contrairement à ce qui a été obtenu en partie 5.2.1.1, avec un maximum de bonne reconnaissance obtenu avec le bruit ♯2 et l'utilisation du PCE moyen pour la classification de $83{,}34\%$. Dans l'expérimentation précédente, il s'agissait pour chaque filtre de la base de définir, pour un seuil donné, s'il s'agissait de la même personne. Ainsi, nous pouvions rencontrer des cas où l'image cible était affectée à aucune ou à plusieurs personnes différentes de la base d'apprentissage. Dans l'expérimentation présentée ici, il s'agit d'affecter à chaque image cible un unique individu de la base d'apprentissage. En outre, une tentative de génération automatique des régresseurs du bruit a été effectuée ici, alors qu'ils étaient choisis manuellement précédemment. Finalement, nous observons des résultats extrêmement variables en fonction du modèle de bruit choisi. Alors que le taux de bonne reconnaissance avec le PCE moyen est de $83{,}34\%$ pour le bruit ♯2 il est de $59{,}52\%$ pour le bruit ♯1. La sélection des régresseurs du bruit semble donc être une étape critique et une limitation de notre approche.

TABLE 5.4 – Comparaison des méthodes de reconnaissance : pourcentage de bonne et de mauvaise reconnaissance pour les méthodes Eigenfaces, Fisherfaces, LBPH, VLC et VLC avec débruitage du plan de corrélation (VLC+LM).

Méthode				Bonnes détections(%)	Fausses alarmes(%)
Eigenfaces				83,34	16,67
Fisherfaces				95,24	4,76
LBPH				95,24	4,76
VLC			PCE maximum	92,86	7,14
			PCE moyen	97,62	2,38
	VLC + LM	Bruit ♯1	PCE maximum	54,76	45,24
			PCE moyen	59,52	40,48
		Bruit ♯2	PCE maximum	59,52	40,48
			PCE moyen	83,34	16,67

Les résultats obtenus à l'aide du modèle de bruit ♯3, pour lequel les régresseurs sont sélectionnés aléatoirement, sont présentés dans le tableau 5.5 pour six tirages aléatoires sans remise. Les conditions expérimentales sont similaires à celles utilisées pour la génération des résultats précédents. La classification des résultats a été effectuée à la fois avec le maximum de PCE et la moyenne de PCE pour l'ensemble des filtres de chaque personne de la base d'apprentissage. Une grande variabilité du taux de bonne reconnaissance est observable, de l'ordre de $10pts$ pour les deux méthodes de classification. Cela corobore notre hypothèse précédente, à savoir que les résultats sont extrêmement dépendant du choix des régresseurs du bruit, rendant notre algorithme relativement instable. Enfin, nous observons un taux de bonne reconnaissance significativement plus élevé avec la classification à partir du PCE moyen, de l'ordre de $20pts$.

TABLE 5.5 – Pourcentage de bonnes et de mauvaises reconnaissances pour la méthode VLC avec débruitage du plan de corrélation (VLC+LM), à partir d'un bruit aléatoire de 5 images par personne de la base d'entrainement et par filtre., suivant la méthode de classification des résultats (le maximum du PCE ou sa moyenne pour chaque filtre de la personne) pour 6 tirages aléatoires.

Classification	Tirage	Bonnes détections(%)	Fausses alarmes(%)
PCE maximum	♯1	61,90	38,10
	♯2	64,29	35,71
	♯3	66,67	33,33
	♯4	69,05	30,95
	♯5	69,05	30,95
	♯6	59,52	40,48
PCE moyen	♯1	80,95	19,05
	♯2	88,10	11,90
	♯3	85,71	14,29
	♯4	85,71	14,29
	♯5	76,19	23,81
	♯6	78,57	21,43

5.2.3 Conclusion

Notre méthode d'identification VLC avec un débruitage du plan de corrélation par décomposition en modèle linéaire est potentiellement à même d'améliorer significativement les performances de reconnaissance. Malheureusement, celle-ci est largement dépendante du choix des régresseurs du bruit et du signal. Les régresseurs du bruit ne pouvant être générés automatiquement, une étape de sélection à partir de plans de corrélation réels est nécessaire. Comme nous l'avons vu, celle-ci est relativement efficace lorsqu'elle est effectuée manuellement. Malheureusement, les tentatives de sélection automatique que nous avons expérimenté ne sont pas pertinentes, dégradant les résultats obtenus avec la corrélation VLC seule. La réalisation d'une étape de sélection automatique des régresseurs du bruit nécessite des travaux supplémentaires. Enfin, la corrélation VLC dans nos conditions d'utilisation permet l'obtention de résultats comparables aux approches témoins avec lesquelles nous l'avons comparée, à savoir Eigenfaces, Fisherfaces et LBPH. Les résultats sont à nuancer par le faible nombre d'images présentes dans la base.

5.3 Algorithme de suivi et détection de la chute

Nous proposons une approche basée sur le suivi du visage de la personne. Ce choix a été pris car il s'agit de la partie du corps subissant la plus grande accélération verticale lors d'une chute, la seule qui ne soit pas recouverte par des vêtements et enfin la plus caractéristique visuellement d'une personne donnée. Le visage est donc particulièrement adapté au suivi par caméra vidéo. Enfin, et ce pour permettre un temps de calcul faible et pour réduire les coûts d'installation à large échelle, nous proposons l'utilisation de caméras de faible résolution, de type webcam.

Notre algorithme est basé sur un système de suivi itératif. Le corrélateur à spectre joint (JTC) pose les fondements de notre méthode. Il est complété par une méthode de détection de la perte du suivi. Nous introduisons tout d'abord un bref résumé de notre approche de suivi. Dans un second temps nous présentons notre protocole d'expérimentation permettant d'évaluer les performances du suivi lors d'événements variés dans la séquence vidéo. Enfin nous présentons et évaluons un critère de détection des chutes adapté au suivi mono-caméra (sans prise en compte de la profondeur).

5.3.1 Méthode de suivi de visages

Notre algorithme (Fig. 5.5), présenté au chapitre 4, utilise le principe itératif appliqué au corrélateur à spectre joint pour effectuer un suivi de la tête de la personne : la région détectée au temps t est utilisée comme référence du corrélateur au temps $t + 1$. La détection de la région correspondant à l'image référence dans l'image cible est assurée par le JTC, dont les pics d'inter-corrélation sont positionnés suivant la position relative de l'image référence et de sa correspondance. L'initialisation du système est rendu possible par l'application d'une méthode de détection de visages, tel le classifieur de P. Viola et M. Jones [122]. Comme nous l'avons vu dans le chapitre 4, l'emploi de la mesure PCE et la mise en place d'un seuil en deçà duquel l'objet recherché est considéré comme absent de l'image cible ne sont pas adaptés à cette application. Ainsi, il n'est pas possible d'écarter une image de la séquence pour laquelle la corrélation nous renvoie une région aberrante de l'image cible. Cette lacune est particulièrement problématique dans le cas d'une méthode itérative, la région détectée étant dans tous les cas utilisée pour l'itération suivante comme référence. L'objet suivi est par conséquent perdu.

Pour remédier à cela nous avons proposé deux optimisations du système, l'une basée sur la délimitation d'une région d'intérêt dans l'image cible, l'autre sur une étape supplémentaire de validation de la région détectée. La première optimisation, dont l'effet de la taille a été expérimenté au chapitre 4, a permis une amélioration significative du suivi, limitant la région de recherche et donc le nombre de perturbations extérieures induites par le fond de l'image. La quantité de cas où l'objet suivi est perdu a donc été réduit. La seconde optimisation consiste a comparer les histogrammes en niveau de gris de l'image référence et de la région détectée par le corrélateur. Cette étape, très peu coûteuse en calcul, autorise la détection des cas de mauvaise corrélation et donc la réinitialisation de notre méthode.

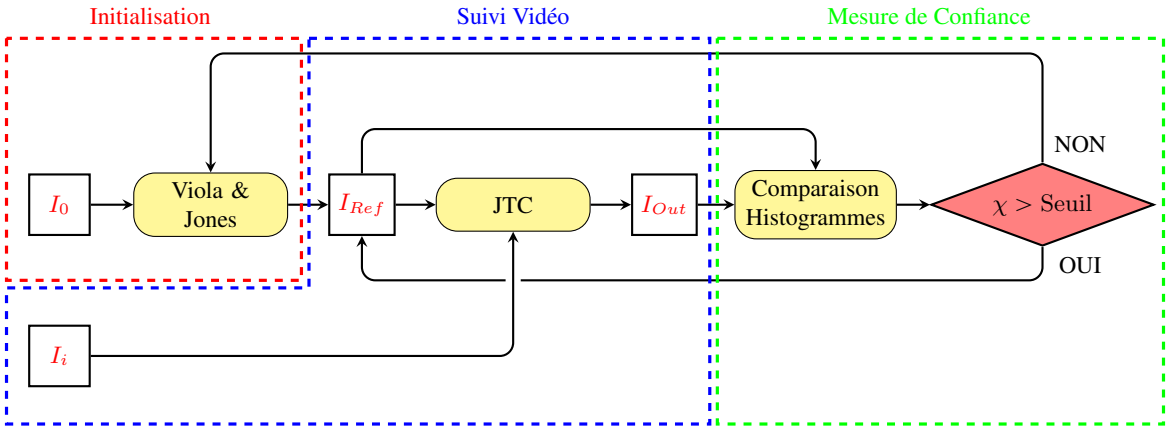


FIGURE 5.5 – Synopsis de l'algorithme itératif par corrélation (Joint Tranform Correlator) avec correction par histogrammes. I_0 correspond à l'image d'initialisation, I_i à la $i^{ème}$ image de la séquence vidéo, I_{Ref} et I_{Out} à l'image extraite de I_i et χ le résultat de notre mesure de similarité d'histogrammes (comparée à un seuil).

Afin d'expérimenter en profondeur notre algorithme, nous définissons un protocole expérimental explorant les limites d'une méthode de suivi. Pour ce faire, nous avons imaginé une large variété de scénarios. Les prises de vues ont été réalisées dans notre reproduction de chambre de centre de prise en charge de personnes dépendantes. La base de données de séquences vidéo ainsi enregistrée contient 25 événements différents afin d'isoler ceux entraînant notre algorithme dans ses conditions limites et afin de les classer par situation. Les différents événements sont présentés sur le tableau 5.6. Ces événements sont organisés par type de mouvement (i.e. position orthostatique, rotation du visage, translation, chute, occlusion ou sortie de champ), par vitesse d'exécution (i.e. rapide ou lente), et par direction (i.e. haut, bas, latéral). Les événements expérimentant le suivi

sont affichés en bleu, ceux expérimentant la chute, en rouge. Certains de ces événements sont illustrés en figure 5.6 : la translation verticale (Fig. 5.6a et 5.6b), la rotation circulaire du visage (Fig. 5.6c) et la chute frontale (Fig. 5.6d). Ils correspondent respectivement aux événements ♯16, ♯6 et ♯21 sur le tableau 5.6. Enfin, chaque événement est précédé et suivi d'une phase d'orthostatisme, à un point fixé dans la pièce d'expérimentation (à $150cm$ de la caméra, pour disposer de séquences les plus proches possibles pour chacune des personnes. La base de donnée est composée en tout de 102 988 images, enregistrées à partir de 11 individus d'origines ethniques et de sexe différents.

TABLE 5.6 – Base de données : description des événements expérimentés.

Evénement					Nombre d'images
#	Mouvement	Vitesse	Direction	Distance (cm)	
1	Orthostatisme			150	40 924
2	Rotation du visage	Lente	Horizontale	150	6 332
3	Rotation du visage	Rapide	Horizontale	150	2 166
4	Rotation du visage	Lente	Verticale	150	4 690
5	Rotation du visage	Rapide	Verticale	150	2256
6	Rotation du visage	Lente	Circulaire	150	4 676
7	Rotation du visage	Rapide	Circulaire	150	2 678
8	Translation	Lente	Arrière		4 738
9	Translation	Rapide	Arrière		2 586
10	Translation	Lente	Avant		4 192
11	Translation	Rapide	Avant		2 172
12	Translation	Rapide	Latérale	150	2 752
13	Translation	Rapide	Latérale	150	2 972
14	Translation	Lente	Haut		2 144
15	Translation	Rapide	Haut		1 164
16	Translation	Lente	Bas		2 500
17	Translation	Rapide	Bas		1 048
18	Occlusions			150	942
19	Sorties de champ				7 898
20	Chute	Lente	Avant	300	1 160
21	Chute	Rapide	Avant	300	520
22	Chute	Lente	Arrière	50	798
23	Chute	Rapide	Arrière	50	464
24	Chute	Lente	Latérale	150	808
25	Chute	Rapide	Latérale	150	418
Total					102 988

Pour déterminer les performances de localisation de notre méthode, il est nécessaire de connaître la position du visage des individus sur chaque image des différentes séquences vidéo. Dans ce but, la position du visage a été enregistrée manuellement par deux opérateurs puis les valeurs obtenues ont été moyennées, nous retournant une "réalité terrain". Également, les événements apparaissant ont été labellisés manuellement par un opérateur.

Les résultats de cette expérimentation apparaissent en tableau 5.7. Y est observable le pourcentage d'images pour lequel le suivi est considéré comme valable pour le suivi avec sans correction d'histogrammes pour des valeurs de décimation allant de 1 à 3 (Partie 4.3.1 - page 89). La mesure de non-similarité par comparaison d'histogrammes (Partie 4.4.2 - page 94) a été calculée pour les deux cas, avec et sans correction. Pour l'algo-

(a) (b) (c) (d)

FIGURE 5.6 – Événements d'expérimentation : (a) et (b) translation vercitale ; (c) rotation circulaire du visage ; (d) chute frontale.

rithme sans optimisation, le visage suivi a été automatiquement considéré comme perdu pour une valeur du Chi Square de Pearson $X^2 > 100$ et lorsque le centre de la région détectée est inclus dans la région obtenue par suivi manuel. Pour ce qui est du suivi JTC avec optimisation par histogrammes, les images pour lesquelles l'individu est détecté à l'aide du JTC ou du classifieur de P. Viola et M. Jones ont été indifféremment considérées comme appartenant à la classe des images avec suivi. Enfin, le suivi avec optimisation a été réinitialisé pour un $X^2 > 100$ également.

Dans un premier temps, nous pouvons observer une importante augmentation du pourcentage d'images pour lesquelles la localisation a été considérée comme valable entre le suivi avec et sans correction par comparaison d'histogrammes. Celle-ci est respectivement, pour des valeurs de décimation de 1,2 et 3, de $49{,}01pts$, $47{,}18pts$ et de $46{,}99pts$. Ainsi, nous sommes en mesure de dire que l'utilisation de l'étape supplémentaire de correction par comparaison d'histogrammes engendre une amélioration significative du processus de suivi. Enfin, nous pouvons observer que le taux le plus élevé d'images dans lequel le visage est correctement détecté est obtenu pour une décimation de $d = 1$.

TABLE 5.7 – Pourcentage d'images suivies en fonction de la décimation et de la présence ou non d'optimisation par comparaison d'histogrammes ($s_{plan} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s \in [78 \times 78px{,}120 \times 120px]$ and $k = 0{,}4$).

	Décimation					
	1		2		3	
Suivi / Valeur	Classique	Histogramme	Classique	Histogramme	Classique	Histogramme
Images suivies (%)	20,51	69,52	18,40	65,58	17,62	64,61
Images suivies (%)	79,49	30,48	81,60	34,42	82,38	35,39

Le tableau 5.8 représente le pourcentage d'images pour lequel le visage a été correctement détecté et positionné pour les 25 événements décrits dans le tableau 5.6, pour un suivi JTC avec et sans optimisation par comparaison d'histogrammes pour des valeurs de décimation de $d = 1$ à $d = 3$. Quelques résultats significatifs peuvent être retenus, spécialement pour les cas limite pour les algorithmes de suivi itératifs, c'est-à-dire les occlusions et les cas de sortie de champ de l'individu (événements ♯18 et ♯19). Tout d'abord nous pouvons observer une amélioration importante du taux de détection (entre $15pts$ et $60pts$ de pourcentage) dans le cas de chute lente et rapide, et cela pour les trois valeurs de décimation et tous les types de chute. Les chutes les moins bien suivies par le système correspondent aux chutes latérales et à la chute rapide avant (environ 55% d'images suivies). Une seconde amélioration significative a lieu pour les translations rapides vers le haut, passant de 0% à $78{,}69\%$ images suivies. Concernant les occlusions et les sorties de champ, le taux de détection présente une augmentation de $11{,}78pts$, $13{,}69pts$ et $15{,}4pts$ pour $d = 1$ à $d = 3$, respectivement.

Notre optimisation du suivi JTC itératif à l'aide de l'introduction d'une mesure de non-similarité des histogrammes introduit une amélioration significative des résultats, réduisant le nombre de cas de non détection du visage entre deux images consécutives. En effet, tandis qu'une perte de la région suivie était auparavant définitive, cette optimisation rend notre algorithme capable de détecter de telles situations et ainsi de se réinitialiser automatiquement. La décimation augmentant le taux de non détection, elle permet, dans le cas de l'algorithme avec optimisation, de réaliser un compromis entres le JTC et le classifieur de P. Viola et M. Jones. Le meilleur compromis pour ces données est obtenu pour une décimation de $d = 1$, autorisant la prise en compte des avantages des deux méthodes.

5.3.2 Critère de détection des chutes

Nous situant dans le cadre d'une détection de chutes des personnes âgées, il est nécessaire de déterminer un critère caractérisant une chute. Comme nous l'avons expliqué précédemment, une chute brutale peut-être définie par le passage dans un faible laps de temps d'une position debout à une position allongée. Ainsi, il est possible à l'aide de la vitesse verticale du visage de définir un seuil à partir duquel une chute est considérée comme ayant eu lieu. Malheureusement, ce type de critère ne prend pas en compte les chutes suivant un mouvement elliptique (de vitesse verticale plus faible). La formule décrite par l'équation 5.1 est basée sur une mesure de la vitesse verticale. Afin de permettre une adaptation du critère selon le type de chute, la vitesse horizontale est également prise en compte. Étant moins importante que la verticalité, une pondération de $1/4$ lui a été expérimentalement appliquée.

$$(y_t - y_{t-1}) + \frac{1}{4} \times |x_t - x_{t-1}| > Seuil \tag{5.1}$$

— *x_t et y_t sont les coordonnées du visage dans l'image au temps t.*

Les résultats de notre critère de détection des chutes sont présentés sur le tableau 5.9. Le seuil définissant la détection d'une chute a été fixée à une variation supérieure à $450px$ dans un laps de temps de $1{,}34s$. Nous pouvons observer tout d'abord un très faible nombre de fausses alarmes. En effet, seules les translations rapides verticales et les occlusions ont généré une alerte indésirable. Également, les chutes arrières sont les mieux détectées par notre système. Cela est explicable par la présence continuelle du visage de face dans l'image durant la chute. Enfin, les autres cas de chute génèrent très peu d'alarmes. Par conséquent, notre critère de détection des chutes est peu performant. Cela est explicable par trois facteurs. Premièrement, la vitesse de la chute mesurée dépend de la distance entre le sujet et la caméra. Deuxièmement, lors d'une perte du suivi du visage durant une chute, le classifieur de P. Viola et M. Jones a une faible probabilité de retrouver le visage de la personne. Enfin notre méthode de détection elle-même n'est pas adaptée pour des chutes lentes et insuffisante pour détecter efficacement une chute suivant une trajectoire elliptique. Nous présentons les performances du suivi et de la détection de chutes en tableau 5.10. Une faible efficacité de la détection de la chute est observable pour les personnes 4, 9 et 10. Cependant, le pourcentages d'images avec un suivi JTC est de l'ordre de la moyenne, au dessus de 60% pour les personnes 4 et 9. Le faible nombre de chutes détectées pour ces deux

TABLE 5.8 – Pourcentage d'images suivies pour chaque événement en fonction de la décimation et de la présence ou non d'optimisation par comparaison d'histogrammes pour chaque type d'événement ($s_{plan} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s \in [78 \times 78px{,}120 \times 120px]$ and $k = 0{,}4$).

				Décimation					
				1		2		3	
				Suivi JTC					
#	Mouvement	Vitesse	Direction	Classique	Histogramme	Classique	Histogramme	Classique	Histogramme
1	Orthostatisme			100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
2	Rotation du visage	Lente	Horizontale	18,84%	50,28%	12,49%	39,40%	5,69%	35,20%
3	Rotation du visage	Rapide	Horizontale	6,37%	35,09%	4,43%	33,52%	4,16%	33,75%
4	Rotation du visage	Lente	Verticale	6,61%	65,29%	3,05%	52,71%	1,75%	49,00%
5	Rotation du visage	Rapide	Verticale	6,83%	48,89%	3,24%	34,71%	2,39%	43,00%
6	Rotation du visage	Lente	Circulaire	5,88%	21,32%	3,42%	16,72%	3,38%	16,66%
7	Rotation du visage	Rapide	Circulaire	2,28%	16,17%	2,09%	16,13%	1,87%	16,36%
8	Translation	Lente	Arrière	12,85%	87,80%	11,38%	83,79%	9,08%	77,67%
9	Translation	Rapide	Arrière	14,46%	84,53%	8,78%	80,32%	8,35%	79,16%
10	Translation	Lente	Avant	8,23%	87,33%	7,23%	82,23%	5,30%	79,48%
11	Translation	Rapide	Avant	10,41%	86,23%	7,04%	80,89%	7,55%	77,07%
12	Translation	Rapide	Latérale	10,43%	62,06%	5,56%	59,85%	6,94%	59,99%
13	Translation	Rapide	Latérale	13,46%	68,88%	11,47%	65,92%	10,73%	65,81%
14	Translation	Lente	Haut	5,50%	79,34%	1,07%	73,83%	0,00%	74,58%
15	Translation	Rapide	Haut	0,00%	78,69%	0,00%	73,88%	0,77%	72,77%
16	Translation	Lente	Bas	25,12%	82,40%	14,28%	78,20%	8,24%	77,20%
17	Translation	Rapide	Bas	14,98%	77,10%	12,60%	76,34%	11,83%	75,57%
18	Occlusions			0,85%	12,63%	0,85%	14,54%	1,80%	17,20%
19	Sorties de champ			0,00%	0,49%	0,00%	0,51%	0,00%	0,63%
20	Chute	Lente	Avant	30,26%	64,91%	16,29%	52,41%	14,05%	45,43%
21	Chute	Rapide	Avant	27,69%	55,19%	17,31%	46,92%	16,92%	50,00%
22	Chute	Lente	Arrière	12,53%	71,43%	6,39%	70,80%	7,77%	71,18%
23	Chute	Rapide	Arrière	6,68%	68,10%	4,74%	65,73%	5,39%	69,83%
24	Chute	Lente	Latérale	37,13%	52,72%	21,78%	46,91%	17,70%	44,80%
25	Chute	Rapide	Latérale	33,97%	55,98%	32,06%	55,98%	29,67%	58,61%

personnes s'explique par le nombre élevé de perte de suivi de la tête lors des événements "chute" en particulier. En effet, il s'agit des événements pendant lesquels le mouvement est le plus rapide, engendrant un flou sur l'image. La tête est donc plus susceptible d'être perdue par l'algorithme. De plus, les caméras sont équipées d'un système autofocus qui n'est pas suffisamment performant lors d'événements rapides, augmentant d'autant le flou présent sur l'image. Pour ce qui est de la personne 10, le nombre de chutes détectées est dû au faible nombre d'images dans lesquelles le visage est correctement localisé. En effet, la détection du visage lors de l'initialisation à l'aide du classifieur de P. Viola et M. Jones est peu efficient pour cette personne précise pour des raison de réflexion de la lumière sur haut du crâne.

5.3.3 Conclusion

L'utilisation d'une mesure de non-similarité des histogrammes permet une amélioration significative des performances du suivi par JTC, nous permettant d'atteindre un taux de $69{,}52\%$ d'images avec un suivi efficace. L'étape de détection de chutes correspond à une première expérimentation est nécessite des travaux plus étendus. Elle permet d'obtenir 25% de bonne détection pour $4{,}37\%$ de fausses alarmes, ce qui représente 33 chutes détectées sur les 132 chutes exécutées. Les déficiences de cette étape sont dues à la fois à l'étape de suivi, et au critère de détection de la chute utilisé. En effet, différentes lacunes de notre algorithme persistent pour une telle application. Premièrement, lors d'une perte de suivi au cours de la chute, l'étape d'initialisation du suivi ne permet pas d'obtenir un taux de détection du visage acceptable. De plus, des mouvement rapides engendrent un flou sur l'image capturée, perturbant la corrélation. Ces deux effets combinés augmentent à la fois la probabilité de décrochage du système tout en réduisant sa capacité à se réinitialiser. Deuxièmement, le critère de détection n'est adapté qu'aux chutes brutales. Les chutes molles ou syncopales (le sujet se retient à un meuble ou perd connaissance) peuvent être détectées que par une méthode décision beaucoup plus développée, basée sur une analyse approfondie de la chute. Finalement, des déplacements dans la profondeur influent à la fois sur le critère de détection et sur le JTC. La corrélation étant très sensible au changement d'échelle, la probabilité de perte de suivi en est d'autant plus augmentée. En outre, notre critère de détection ne peut prendre en compte des chutes dans cette direction.

5.4 Perspectives

Comme nous venons de le voir, l'utilisation d'un système monoscopique ne permet pas un suivi suffisamment performant de la personne pour une détection efficace des chutes d'un individu.

Tout d'abord, l'absence d'information de profondeur réduit considérablement la pertinence de notre critère de détection des chutes. En effet, celui-ci est basé sur la vitesse du visage. En l'absence de l'information de profondeur, la seule mesure de vitesse que l'on peut effectuer est issue de la distance en pixels sur l'image. Or, celle-ci ne peut être valable uniquement pour une distance donnée par rapport à la caméra. Effectivement, comme nous le présentons en figure 5.7, une même hauteur dans la réalité ne donne pas une même distance en pixel que le sujet soit proche ou éloigné de la caméra. Un sujet proche (Fig. 5.7a) chutera d'une hauteur h_1 plus grande en pixels sur une image qu'un sujet éloigné (Fig. 5.7b). Ainsi, pour être en mesure de mettre en place un critère de détection de chutes basé sur la vitesse du visage de la personne, nous devons impérativement être en possession des distances réelles, en mètre, et donc disposer des informations de profondeur.

En outre, la région délimitant le visage conserve la même taille tout au long de la séquence vidéo, que la personne soit proche ou non de la caméra vidéo. Ceci engendre tout d'abord une augmentation de bruit dans l'image : au fur-et-à-mesure que le sujet s'éloigne, une partie de plus en plus grande du fond de l'image se retrouve inclue dans cette région, rendant le système plus instable. En effet, la probabilité que le JTC corrèle avec le fond de l'image augmente dans ces conditions. À l'inverse, si la personne se rapproche de la caméra, une plus faible portion du visage est inclue dans la région le délimitant, entraînant une perte d'information. En outre, la région d'intérêt, dans laquelle le visage est recherché dans l'image cible, explicitée en partie 4.4.1,

TABLE 5.9 – Pourcentage de chutes détectées pour chaque événement en fonction de la décimation et de la présence ou non d'optimisation par comparaison d'histogrammes ($s_{plan} = (512{,}512)px$, $s_{roi} = (3 \times s)px$, $s \in [78 \times 78px{,}120 \times 120px]$ and $k = 0{,}4$).

				Décimation					
				1		2		3	
				Suivi JTC					
#	Mouvement	Vitesse	Direction	Classique	Histogramme	Classique	Histogramme	Classique	Histogramme
1	Orthostatisme			0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	Rotation du visage	Lente	Horizontale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
3	Rotation du visage	Rapide	Horizontale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4	Rotation du visage	Lente	Verticale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
5	Rotation du visage	Rapide	Verticale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
6	Rotation du visage	Lente	Circulaire	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
7	Rotation du visage	Rapide	Circulaire	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
8	Translation	Lente	Arrière	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
9	Translation	Rapide	Arrière	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
10	Translation	Lente	Avant	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
11	Translation	Rapide	Avant	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
12	Translation	Rapide	Latérale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
13	Translation	Rapide	Latérale	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
14	Translation	Lente	Haut	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
15	Translation	Rapide	Haut	11,36%	11,36%	11,36%	11,36%	11,36%	11,36%
16	Translation	Lente	Bas	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
17	Translation	Rapide	Bas	13,64%	13,64%	13,64%	13,64%	13,64%	13,64%
18	Occlusions			0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
19	Sorties de champ			46,10%	46,10%	46,10%	46,10%	46,10%	46,10%
20	Chute	Lente	Avant	4,55%	4,55%	4,55%	4,55%	4,55%	4,55%
21	Chute	Rapide	Avant	4,55%	4,55%	4,55%	4,55%	4,55%	4,55%
22	Chute	Lente	Arrière	27,27%	27,27%	27,27%	27,27%	27,27%	27,27%
23	Chute	Rapide	Arrière	86,36%	86,36%	86,36%	86,36%	86,36%	86,36%
24	Chute	Lente	Latérale	18,18%	18,18%	18,18%	18,18%	18,18%	18,18%
25	Chute	Rapide	Latérale	9,09%	9,09%	9,09%	9,09%	9,09%	9,09%

TABLE 5.10 – Performances de l'algorithme de suivi (avec correction par comparaison d'histogrammes) et de la détection des chutes pour chacune des 11 personnes de la base de données.

Personne	Performances de suivi (en % d'images)		Détection de chutes	
1	JTC	69,89%	Fausses Alarmes	7,65%
	VLC	3,83%	Bonnes détections	16,67%
	Images non suivies	26,28%		
2	JTC	19,48%	Fausses Alarmes	7,41%
	VLC	4,99%	Bonnes détections	25,00%
	Images non suivies	75,54%		
3	JTC	48,67%	Fausses Alarmes	5,88%
	VLC	7,66%	Bonnes détections	20,00%
	Images non suivies	43,68%		
4	JTC	67,67%	Fausses Alarmes	6,88%
	VLC	5,44%	Bonnes détections	0,00%
	Images non suivies	26,90%		
5	JTC	69,56%	Fausses Alarmes	3,57%
	VLC	11,70%	Bonnes détections	60,00%
	Images non suivies	18,74%		
6	JTC	59,21%	Fausses Alarmes	4,71%
	VLC	8,47%	Bonnes détections	33,33%
	Images non suivies	32,32%		
7	JTC	68,53%	Fausses Alarmes	3,49%
	VLC	7,47%	Bonnes détections	41,67%
	Images non suivies	23,99%		
8	JTC	66,00%	Fausses Alarmes	3,45%
	VLC	4,86%	Bonnes détections	16,67%
	Images non suivies	29,14%		
9	JTC	69,32%	Fausses Alarmes	3,89%
	VLC	7,99%	Bonnes détections	14,29%
	Images non suivies	22,70%		
10	JTC	28,16%	Fausses Alarmes	1,76%
	VLC	4,74%	Bonnes détections	8,33%
	Images non suivies	67,10%		
11	JTC	68,43%	Fausses Alarmes	6,11%
	VLC	13,51%	Bonnes détections	50,00%
	Images non suivies	18,06%		

page 93, est définie par un facteur appliqué à la région délimitant le visage. Ce facteur est fixé en fonction de la probabilité que le visage se retrouve dans cette région dans l'image suivante. Si la personne s'approche de la caméra vidéo, le risque que la région d'intérêt soit trop restreinte, c'est-à-dire que le visage de l'individu se trouve en dehors de cette zone, augmente. À l'opposé, celle-ci devient trop grande lorsque la personne s'éloigne, augmentant encore une fois la sensibilité de l'algorithme au bruit induit par le fond de l'image.

Finalement, la connaissance de la position de la personne dans l'espace autorise la mise en place de zones spécifiques dans la pièce dans lesquelles les critères d'émission de signaux d'alertes pourront être modifiés, comme un lit, dans lequel l'individu pourra évidemment être en décubitus, ou un fauteuil, où l'inaction de la personne pendant un laps de temps relativement long pourra constituer une situation d'alerte.

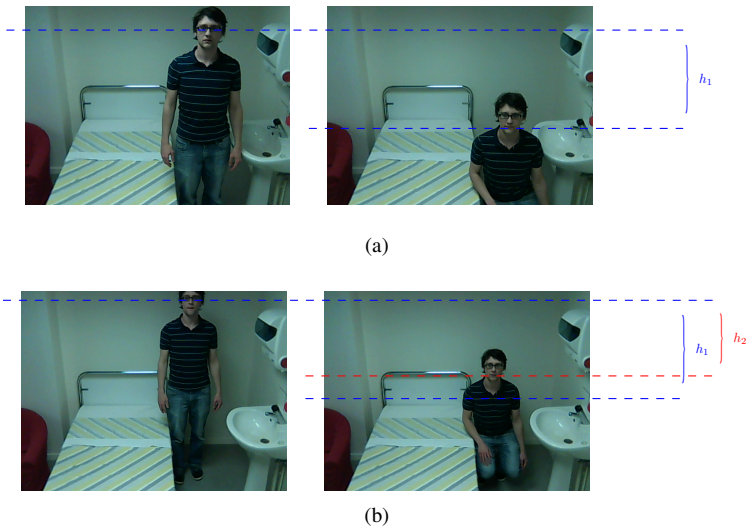


FIGURE 5.7 – Nécessité de la connaissance de la profondeur : (a) hauteur de chute pour une personne proche de la caméra ; (b) hauteur de chute pour une personne éloignée de la caméra.

Déterminer l'information de profondeur est impossible au moyen d'une unique caméra fixe. Pour résoudre un tel problème, il est nécessaire de disposer soit d'un système comprenant plusieurs caméras, soit d'une caméra se déplaçant autour d'un objet fixe. Nous situant dans le cadre d'une méthode de suivi, l'utilisation d'une caméra mobile est écarté. Nous préférerons la prise en compte de données issues d'un système bi-caméra, la stéréovision.

Le principe de détermination de la position 3D d'un point à partir de deux caméras réside dans le fait que ce même point dans l'espace est visible à partir de deux vues différentes. Cette approximation, la triangulation, réalisée à partir de méthodes comme la DLT ou l'approximation de Sampson, n'est envisageable que si on est possession de la relation existante entre les deux caméras, les caractéristiques extrinsèques, et des transformations propres à chaque caméra, les caractéristiques intrinsèques. L'étape de calcul de ces points, appelée étape de calibrage, est effectuée à l'aide de points dont les positions relatives sont connues, par exemple un damier.

Ainsi, pour être en mesure de connaître la position d'un point dans les trois dimensions à partir de deux caméras calibrées (dont les caractéristiques extrinsèques et intrinsèques sont connues), il est nécessaire de disposer de la position de ce point en deux dimensions sur chacune des caméra. Il existe deux approches pour réaliser cette étape. La première est de rechercher indépendamment, sur chacune des caméras, le point d'intérêt. Nous nous situons dans le cadre de l'application d'un suivi itératif basé sur la corrélation. Il s'agit donc à partir de deux images de référence différentes, une pour chaque caméra, de retrouver un même point d'intérêt. De plus, ce point doit être un point particulier de la région recherchée (par exemple le barycentre). Il est donc très peu probable, dans ces conditions, d'obtenir exactement le même point sur chacune des deux images. Enfin, du fait de l'utilisation de transformées de Fourier directes et inverses sur un plan d'entrée volumineux, la corrélation JTC est coûteuse en temps de calcul. Une telle approche n'est donc pas envisageable pour notre application ; La seconde méthode consiste à prendre en considération les relations existantes entre les deux images pour être en mesure d'utiliser ce que l'on appelle les lignes épipolaires. Une ligne épipolaire, schématisée en figure 5.8, est définie par l'intersection entre le plan épipolaire (le plan passant par le point d'intérêt et la ligne joignant les centres optiques des caméras vidéo) et les images issues de chacune des deux caméras.

Connaissant les relations de transformation entre les deux caméras, nous sommes donc en mesure de calculer la ligne épipolaire pour un point particulier d'une des deux caméras. Cette ligne étant l'intersection

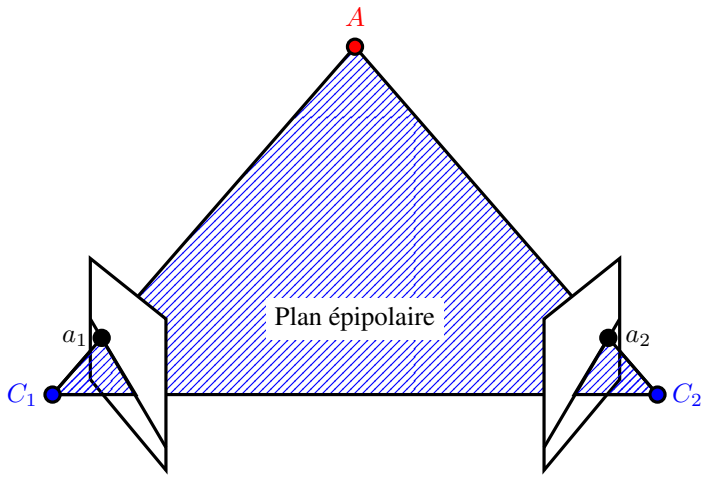


FIGURE 5.8 – Géométrie épipolaire. Les points a_1 et a_2 correspondent à la projection du point A sur les images 1 et 2 issues de chacune des caméras du dispositif, représentées par leurs centres optiques C_1 et C_2.

du plan épipolaire avec les images pour chacune des deux caméras, un point dans une des deux caméra sera donc situé sur la ligne épipolaire de l'autre caméra. Conséquemment, il suffit de rechercher son correspondant sur cette ligne. Pour ce faire nous utilisons à nouveau la comparaison des histogrammes de la région détectée dans l'image de la caméra 1 avec toutes les régions de mêmes dimensions dont le barycentre est sur la ligne épipolaire. Afin de réduire la probabilité de mauvaise mise en correspondance, l'histogramme n'est pas calculé sur la région entière. Celle-ci est fractionnée en 4 sous-parties. Sur chacune de ces sous-parties l'histogramme est calculé et comparé avec sa sous-partie correspondante dans la région de référence à l'aide du Chi Square de Pearson. Finalement, la valeur de non-similarité entre les deux régions correspond à la somme des quatre mesures calculées. La région conservée est celle obtenant la valeur la plus faible.

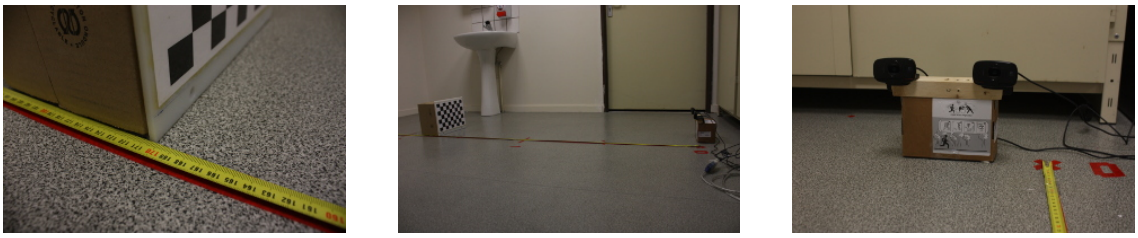
Notre de méthode de suivi stéréoscopique est basée sur notre algorithme monoscopique de suivi JTC sur une des caméras vidéo, la position du visage sur la seconde caméra est détectée en utilisant la mise en correspondance à l'aide de la géométrie épipolaire. L'étape d'initialisation est effectuée simultanément sur les deux caméras, à l'aide du classifieur de P. Viola et M. Jones [122]. Lorsqu'un visage est détecté sur une image de l'une des deux caméras, notre algorithme de suivi par JTC, présenté au chapitre 4 et en partie 5.3, page 112, y est appliqué. À chaque laps de temps, la région correspondant à celle détectée à l'aide du JTC est déterminée sur la ligne épipolaire comme énoncé précédemment. À partir des deux barycentres des régions obtenues sur chacune des deux images, la position tridimensionnelle du visage est calculée par triangulation (avec la méthode DLT). Finalement, les dimensions de la région d'intérêt, dans laquelle le visage est recherché dans l'image cible à l'aide de l'algorithme JTC, est mise-à-jour à partir des données de profondeur obtenues. En cas de détection d'une perte de la région suivie à l'aide de l'optimisation du suivi JTC par comparaison des histogrammes, l'algorithme est réinitialisé.

Notre dispositif de suivi stéréoscopique est présenté en figure 5.9. Il s'agit de deux webcams Logitech HC584. Afin de déterminer la précision de notre système de stéréovision et de choisir entre différentes configurations de calibration, nous avons mis en place un protocole d'expérimentation. Le système stéréoscopique est positionné sur une graduation. La mire (le damier utilisé pour la calibration) est déplacée le long de cette graduation avec un pas de $100cm$. 26 images sont ainsi acquises. Ce protocole est présenté en figure 5.10.

Le tableau 5.11 permet de visualiser numériquement la distance estimée par le système en fonction de la distance mesurée manuellement pour une calibration utilisant 1 à 26 images avec un pas de $10cm$. On observe une erreur moyenne et quadratique plus faible ($22{,}04$ et $22{,}43$ respectivement) pour l'utilisation de 10 paires

(a) (b) (c)

FIGURE 5.9 – Système d'acquisition : (a) les deux caméras ; (b) le dispositif complet avec son trépied ; (c) le système en situation.



(a) (b) (c)

FIGURE 5.10 – Protocole de validation de la calibration : (a) la mire placée sur la graduation ; (b) le dispositif complet ; (c) le système stéréoscopique.

d'images (soit des images éloignées de maximum $140cm$ du dispositif d'acquisition). L'erreur de reprojection minimale obtenue expérimentalement apparaît donc pour une utilisation de 10 images de calibration par caméra. L'erreur théorique optimale quant à elle est estimée à 8 paires d'images. On utilisera donc un nombre compris entre 8 et 10 images de calibration, distantes de moins de $140cm$ du système.

Finalement nous avons appliqué notre méthode de suivi stéréoscopique sur une séquence de tests (1390 images), ceci afin de déterminer à quel point il affectait les résultats par rapport au suivi mono-caméra. La séquence utilisée contient plusieurs chutes, des rotations du visage, rotation du sujet, un saut sur place et un saut vers l'avant (limites connues du système). L'algorithme mono-caméra a été appliqué séparément sur chacune des caméra gauche et droite (L et R). L'algorithme stéréo a également été appliqué et les résultats, affichés sur le tableau 5.12, ont été illustrés pour chacune des caméras séparément (L et R) et pour le système dans son ensemble. On observe que le système est peu affecté par le suivi stéréo Mais présente l'intérêt, majeur pour nous, de permettre une estimation de la position 3D.

5.5 Conclusion

Dans ce chapitre, nous avons présenté une application de la corrélation à un système de détection des chutes de la personne dépendante. L'objectif de l'approche était d'être en mesure de détecter efficacement une chute brutale, tout en s'affranchissant des contraintes inhérentes au port de capteurs par l'individu. De plus, le dispositif devait être à même de pouvoir être installé dans l'habitat de la personne ou dans des centres de prise en charge de personnes dépendantes sans nécessiter de travaux importants et de coûts élevés, ce afin de

TABLE 5.11 – Evaluation de la calibration pour différents nombres de paires d'images de calibration. La première partie du tableau représente la distance mesurée. La seconde partie représente la différence entre la distance théorique et la distance mesurée. La dernière partie contient la moyenne quadratiques et arithmétiques (RMSE) des erreurs.

Distance	11	10	9	8	7	6	5	4
Distance mesurée								
50	35,79	68,08	69,92	88,50	82,84	78,60	76,44	77,94
100	63,90	122,11	125,38	158,91	148,96	141,64	138,30	141,39
150	88,57	169,74	174,22	220,94	207,38	197,60	193,64	198,38
200	114,72	220,32	226,05	286,75	269,53	257,38	253,17	259,93
250	145,57	279,94	287,07	364,45	343,04	328,38	324,52	334,05
Différence entre la distance mesurée et la distance théorique								
50	14,21	18,08	19,92	38,50	32,84	28,60	26,44	27,94
100	36,10	22,11	25,38	58,91	48,96	41,64	38,30	41,39
150	61,43	19,74	24,22	70,94	57,38	47,60	43,64	48,38
200	85,28	20,32	26,05	86,75	69,53	57,38	53,17	59,93
250	104,43	29,94	37,07	114,45	93,04	78,38	74,52	84,05
Moyennes quadratiques et arithmétiques des erreurs								
Moyenne	60,29	22,04	26,53	73,91	60,35	50,72	47,21	52,34
RMSE	68,49	22,43	27,13	78,24	63,65	53,39	49,90	55,66

TABLE 5.12 – Analyse de 1390 images à l'aide de l'algorithme JTC, pour le système stéréovision et le système monovision.

	Stéréovision			Monovision	
Distance moyenne (px)	41,428	18,633	18,595	16,973	18,595
♯ décrochages	11	11	11	11	9
♯ images suivies (JTC)	0	1319	1319	1312	1319
♯ images suivies (VJ)	2	10	2	12	10
♯ images suivies (Total)	1331	1330	1329	1327	1328
♯ images non suivies (Total)	59	60	61	63	62

pouvoir être utilisé à large échelle.

Nous avons donc proposé l'utilisation de l'information issue de caméras vidéo de faible résolution, ce afin de respecter la contrainte de coût et d'être à même de traiter l'information dans un temps suffisamment bref pour pouvoir alerter le personnel de secours.

Afin d'éviter un suivi non pertinent d'une personne, nous avons apporté une étape d'identification. Celle-ci est basée sur la corrélation Vander Lugt. Une étape de débruitage, explicitée au chapitre 3, permet d'améliorer significativement les performances et de donner des résultats prometteurs. Cette méthode a été comparée avec des algorithmes standards de la littérature.

Le visage étant la partie du corps détenant le plus d'information visuelle, nous avons choisi de le suivre à l'aide de notre algorithme itératif basé sur le JTC et optimisé à l'aide d'une étape de comparaison des histogrammes consécutifs. On a tout d'abord réalisé un premier système de suivi en deux dimensions, basé sur une seule caméra. Celui apporte des résultats intéressants mais insuffisants pour un suivi robuste du visage et particulièrement pour la détection d'une chute. En effet, l'absence d'information de profondeur rend le système sensible au bruit induit par le fond de l'image et ne permet pas l'utilisation d'un critère efficace de la détection de la chute.

Un second système est en cours de développement, afin d'obtenir l'information de profondeur.

Conclusion et Perspectives

Conclusion et perspectives

Dans ce manuscrit, nous nous sommes focalisés sur la détection des chutes de la personne âgée dépendante à l'aide des méthodes de corrélation. Notre principal objectif, et le besoin formulé par Malakoff-Médérik, était de proposer une méthode ne nécessitant pas le port de dispositif par la personne. Après exploration des différents capteurs utilisables dans une telle application, l'utilisation de l'information vidéo s'est avérée la plus pertinente. Par la suite, l'état de l'art sur les systèmes existant utilisant la vidéo-surveillance pour la détection des chutes nous fait porter notre choix sur l'utilisation d'une méthode de suivi de la personne. En effet, malgré les bonnes performances des systèmes basés sur la posture de la personne, celles-ci ne permettent qu'une analyse partielle de la chute. En outre, elles ne permettent que peu de perspectives d'étendue de la méthode à un système d'habitat intelligent plus global, permettant d'analyser les activités quotidiennes de la personne pour offrir des informations complémentaires aux praticiens. Cependant, les systèmes basés sur le suivi du visage souffrent d'une plus grande instabilité et notamment d'un plus grand taux de fausses alarmes, problématiques pour un système de génération d'alertes. En outre, la prise en compte des données 3D est nécessaire pour disposer d'un système pertinent.

Parmi les méthodes de détection et d'identification notre choix s'est porté sur la corrélation. Cela pour trois raisons. La première est sa capacité à être utilisée aussi bien dans des applications d'identification que de suivi. La seconde est sa capacité d'analyse global de la scène. Enfin, s'agissant des compétences du laboratoire, il nous paraissait pertinent d'évaluer ses performances dans une application réelle. Afin de disposer d'un système faible coût et en temps réel, nous avons pris la direction d'une implantation numérique des méthodes de corrélation. Ceci nous permet de nous affranchir d'une installation optique complexe et coûteuse. En outre, l'amélioration des performances de calcul des systèmes numériques ces dernières années nuance la grande rapidité de calcul de l'architecture optique.

Après une analyse poussée des deux architectures de corrélation présentes dans la littérature, le corrélateur à spectre joint et le corrélateur de Vander Lugt, ainsi que de certaines de leurs optimisation, nous avons décidé d'effectuer l'identification à l'aide du VLC et le suivi avec le JTC. Pour l'identification, cela est motivé par les grandes capacités d'optimisation du filtre de corrélation, permettant la création lors de la phase d'apprentissage d'une kyrielle de filtres adaptés à l'application voulue. Pour le suivi, il ne nécessite pas la création d'un filtre à chaque étape d'un algorithme de suivi itératif, comme cela serait le cas avec un suivi par VLC. De plus, sa grande simplicité d'utilisation pour la localisation de l'image de référence dans l'image cible le rend particulièrement adapté à ce genre de finalité.

Pour ce qui est de l'identification, nous avons proposé une méthode numérique de débruitage du plan de corrélation. L'originalité de cette démarche est qu'elle présente un travail sur le plan de corrélation lui-même alors que les méthodes d'optimisation présentes dans la littérature se focalisent généralement sur le filtre lui-même ou sur le pré-traitement des images de référence et cible. Enfin, nous avons exploré les différents paramètres de notre méthode. Nous avons notamment évalué les conséquences des modifications de la population de régresseurs du signal. Également, nous avons comparé deux fonctions de modélisation du pic de corrélation. L'évaluation de cette méthode sur la base de données PHPID présente des résultats prometteurs, améliorant grandement l'identification par rapport au filtre POF seul. Nous avons dans un second temps développé un algorithme de suivi itératif basé sur le corrélateur à spectre joint non-linéaire sans ordre zéro. L'ensemble des paramètres de notre algorithme ont été explorés à l'aide d'un protocole expérimental que nous

avons imaginé. Outre le choix des paramètres, ces expérimentations nous ont permis de démontrer la capacité d'utilisation de cette méthode pour une application de suivi en temps réel. Finalement, nous avons proposé et validé deux optimisations de notre algorithme, l'une permettant l'utilisation de la connaissance de la position *a priori*, l'autre la ré-initialisation de l'algorithme en cas de perte du suivi, problème inhérent aux algorithmes itératifs.

Pour terminer, nos algorithmes d'identification et de suivi ont été appliqués à un système plus global de détection des chutes de la personne âgée. Pour ce faire nous avons mis en place une chambre d'expérimentation, dans l'objectif de prendre en compte les performances en situation proche de la réalité. Tout d'abord une base de donnée de visages a été réalisée, ce afin d'expérimenter notre méthode d'identification. Notre algorithme a été comparé à des méthodes de la littérature. Celui-ci présente des résultats décevants en conditions réelles. En effet, notre méthode de débruitage dégrade les performances de la corrélation seule. Cela est dû notamment à sa grande dépendance au bruit choisi pour la réalisation du modèle. En effet, malgré nos tentatives, nous n'avons pas été en mesure de déterminer une méthode de sélection ou de modélisation automatique du bruit de corrélation. Cependant, nous avons pu vérifier la pertinence de la corrélation seule quant à ses capacités d'identification. Finalement, nous avons réalisé une base de données comprenant un grand nombre d'événements plausibles lors du suivi d'une personne âgée. Cette base nous a permis de déterminer les performances de suivi de notre algorithme et de sa capacité de détection des chutes. Malheureusement, notre système souffre de l'instabilité intrinsèque des méthodes itératives. En outre, l'utilisation de la vitesse verticale et horizontale pour la détection des chute n'est pertinente qu'en possession de l'information de profondeur.

Les perspectives à court termes de ce travail ont pour objectif d'adapter notre algorithme de suivi pour une prise en compte de la profondeur. Pour ce faire, une initialisation simultanée de la méthode sur les deux caméras est nécessaire. La détection d'un visage sur l'une des caméras devra entraîner l'application sur celle-ci de notre algorithme de suivi itératif. Puis, l'utilisation d'une calibration des caméras devra permettre la localisation sur la ligne épipolaire du visage de la personne dans la seconde caméra. L'idée principale de cette méthode est de réduire le nombre de décrochages et d'accélérer l'étape de détection du visage. Finalement l'utilisation de la stéréoscopie permettra l'évaluation de la distance entre les caméras et la personne suivie. Dans un second temps, il sera nécessaire de fusionner l'étape de suivi et d'identification. En effet, il est nécessaire d'enregistrer la trace de l'objet dès sa détection. Pendant le suivi, la personne devra être identifiée. En cas de correspondance avec la personne souhaitée, le suivi pourra continuer à se dérouler normalement. Dans le cas contraire, celui-ci devra éliminer la trace précédemment enregistrée. L'objectif de cette démarche est d'écarter tout d'abord les cas de détection des objets non-humains. En outre, elle rendra possible une utilisation dans une zone d'habitation, dans laquelle des animaux domestiques ou des enfants en bas âge peuvent cohabiter, et perturber l'étape de détection, tout en n'étant pas en mesure par eux-mêmes d'alerter les personnes compétentes.

À moyen terme, nous envisageons tout d'abord d'explorer les possibilités de modélisation automatique du bruit pour l'étape de débruitage de plan de corrélation lors de l'identification. Pour ce faire, il nous est nécessaire de générer un grand nombre de modèle différents afin d'analyser les caractéristiques des plans choisis qui peuvent améliorer ou détériorer le débruitage. Également, dans le cadre du système de détection de chute en lui-même, il serait souhaitable d'augmenter le nombre de caméras utilisées dans la pièce. Ceci pourra se faire soit en installant plusieurs systèmes stéréoscopiques, soit en effectuant une unique calibration pour les différentes caméras. En outre, pour une utilisation en conditions réelles, il sera nécessaire de développer une méthode de sélection de la caméra utilisée pour le suivi, notamment lors de la transition entre différentes pièces.

Les perspectives à long terme consistent en une fusion de plusieurs méthodes et capteurs. En effet, comme nous l'avons vu, les systèmes de détection des chutes basés sur le suivi de la tête souffrent d'une génération d'un grand nombre de fausses alarmes. La limitation de celles-ci par une étape de détection de posture lors de l'hypothèse d'une personne à terre est une piste à explorer. En outre, l'installation de capteurs infrarouge rendrait le système apte à travailler en condition nocturne. Finalement, chacune des méthode et capteurs prise en compte seule souffrent de limitations intrinsèques. Seule une utilisation multimodale et multi-algorithme peut permettre la réalisation d'une méthode non-intrusive de détection des chutes.

Production Scientifique

Production Scientifique

Ayman Alfalou, Christian Brosseau, **Philippe Katz**, Mohammad S. Alam. “Decision optimization for face recognition based on an alternate correlation plane quantification metric”, *Opt. Lett., OSA,* vol. 37, No. 9 (2012), pp. 1562-1564.

Philippe Katz, Ayman Alfalou, Christian Brosseau, Mohammad S. Alam. “Correlation and Independent Component Analysis based approaches for biometric recognition”, Adamo Quaglia, Calogera M. Epifano (Eds.), *Face Recognition : Methods, Applications and Technology, NOVA Publisher,* (2012), pp. 201-229.

Philippe Katz, Michael Aron, Ayman Alfalou. “Joint Transform Correlation for face tracking : elderly fall detection application”, *Proc. SPIE, Optical Pattern Recognition XXIV*, vol. 8748 (2013), pp. 87480I.

Philippe Katz, Michael Aron, Ayman Alfalou. “Joint Transform Correlation pour le suivi de visages : application à la détection des chutes de la personnes âgée”, *Gretsi XXIV* (2013).

Philippe Katz, Michael Aron, Ayman Alfalou. “A face-tracking system to detect falls in the elderly”, *SPIE Newsroom* (2013).

Decision optimization for face recognition based on an alternate correlation plane quantification metric

A. Alfalou,[1,4] C. Brosseau,[2,*] P. Katz,[1] and M. S. Alam[3]

[1]*ISEN Brest, Département Optoélectronique, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France*

[2]*Université Européenne de Bretagne, Université de Brest, Lab-STICC, CS 93837, 6 avenue Le Gorgeu, 29238 Brest Cedex 3, France*

[3]*Department of Electrical and Computer Engineering University of South Alabama, 6001 USA South Dr., Mobile, AL 36688-0002, USA*

[4]*e-mail: ayman.al-falou@isen.fr*

**Corresponding author: brosseau@univ-brest.fr*

Received December 12, 2011; revised January 31, 2012; accepted February 11, 2012;
posted February 13, 2012 (Doc. ID 159824); published May 2, 2012

We consider a new approach for enhancing the discrimination performance of the VanderLugt correlator. Instead of trying to optimize the correlation filter, or propose a new decision correlation peak detection criterion, we propose herein to denoise the correlation plane before applying the peak-to-correlation energy (PCE) criterion. For that purpose, we use a linear functional model to express a given correlation plane as a linear combination of the correlation peak, noise, and residual components. The correlation peak is modeled using an orthonormalized function and the singular value decomposition method. A set of training correlation planes is then selected to create the correlation noise components. Finally, an optimized correlation plane is reconstructed while discarding the noise components. Independently of the filter correlation used, this technique denoises the correlation plane by lowering the correlation noise magnitude in case of true correlation and decreases the false alarm rate when the target image does not belong to the desired class. Test results are presented, using a composite filter and a face recognition application, to verify the effectiveness of the proposed technique.

OCIS codes: 100.5010, 100.3008.

During the last decade, much work has been done in the realm of image recognition techniques. The correlator block diagram shown in Fig. 1 is one of the most promising approaches for these techniques. This diagram consists of an optical part for the correlation process, a filter part, and a decision part. The second and third parts are generally implemented electronically, whereas the first part is usually implemented using an optical setup. To increase the decisional performance of the correlators, much work has been performed by optimizing the correlation filter and/or choosing a specific decision criterion, e.g. signal-to-noise ratio (SNR), and PCE [1–8]. Based on a new concept of correlation filter, we recently suggested an optimization of the correlation plane for a face recognition application [5].

However, this optimization method is limited by the interpretation of the correlation plane. In addition, processing is carried out offline. Here, we propose to apply a postprocessing step in the correlation plane before using a decision criterion. Our specific aim is to increase the correlation peak characteristics when the test image belongs to a predefined class, and decrease the false alarm rate when the target image does not belong to this class.

In this letter, we achieve this by introducing an alternate correlation plane quantification metric, based on the linear functional model (LFM) [8–9] and singular value decomposition (SVD) method [10]. The main advantage of our method lies in its ability to cancel out the background correlation plane (i.e., to denoise the correlation plane) which has for effect to discriminate the correlation peak. Indeed, we assumed that the signal and noise are mutually independent in the correlation plane (input plane of the proposed algorithm). That is, the correlation plane is considered as a combination of the desired signal and background noise. The proposed technique is simple to implement offline and it produces a higher percentage of true class matches in comparison to traditional techniques; i.e., the true positive rate (TPR) is increased by a factor of five for a false positive rate (FPR) set to 0%. After presenting technical aspects of the LFM—SVD algorithm, we present simulation results that illustrate the performance of the proposed technique using the Pointing Head Pose Image Database (PHPID) [11].

The principle of our algorithm is shown in Fig. 2. At first, the input correlation plane P_c [Fig. 2(a)] is decomposed as a linear combination of different regressors. We model the signal and noise regressors by using respectively a set of false correlation planes Y^{noise} [Fig. 2(b)], and a set of sine cardinal functions with different characteristics Y^{peak} [Fig. 2(c)]. Secondly, these regressors are used to describe the input plane with a linear combination of peak, noise components, and a residual signal. Finally, the output correlation plane is reconstructed without the noise components [Fig. 2(d)]. An important advantage of this method is that it gives a well defined and sharp peak for a true correlation and a negligible peak for false correlation.

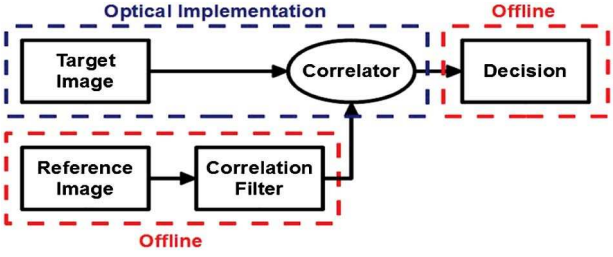


Fig. 1. (Color online) Illustrating the principle of correlation.

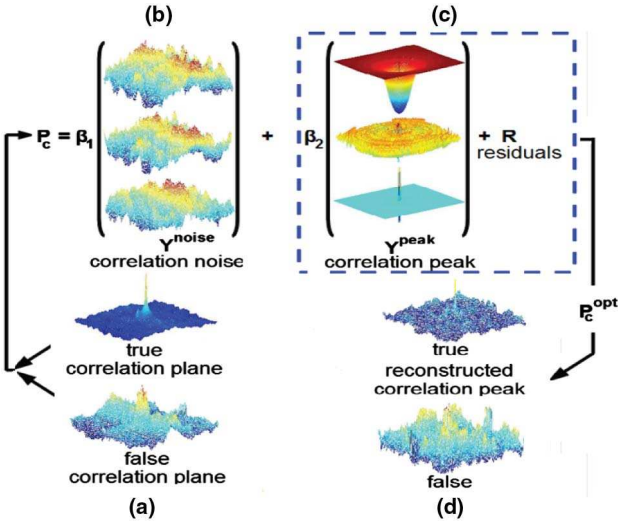


Fig. 2. (Color online) Synoptic diagram of the correlation plane denoising process.

To put this more precisely, the starting point is to set the correlation plane P_c equal to a finite linear combination of weighted regressors, expressed as

$$P_c = \sum_{i=1}^{M} \beta_i Y_i + R, \tag{1}$$

where M is a given integer number, β_i denotes the weight corresponding to regressor $Y_i = Y_i^{\text{noise}} + Y_i^{\text{peak}}$, and R is a residual signal. In practice, it is convenient to rewrite Eq. (1) in a matrix form, given by $P_c = Y\beta + R$, which defines the optimal unbiased estimator β assuming that the residual noise is white, and using the Moore-Penrose pseudoinverse of $\boldsymbol{Y}$. Eq. (1) shows that the decomposition of the correlation plane depends on the regressors; i.e., a sum of weighted regressors is generated to ensure the best possible fit of the real correlation plane. Thus, the input correlation plane (Fig. 2) resulting from the correlation process, called P_c is decomposed into a linear combination of regressors; i.e., correlation noise, correlation peak components, and a residual signal, respectively Y^{noise}, Y^{peak}, and R. As the aim of our algorithm is to reduce the background correlation plane in order to obtain sharper peaks for a match, the signal is then reconstructed by totally discarding the noise components used in the decomposition process, thereby resulting in an optimized correlation plane, P_c^{opt}. The estimation problem amounts to finding specific regressors for both signal and noise [9]. Denoising the correlation plane [P_c^{opt}, Fig. 2(d)] consists in retaining only the desired information (information regressors Y^{peak}) and removing noise information (noise regressors Y^{noise} Fig. 2). Prior to defining the regressors for the information signal, we approximate the correlation peak by a three-dimensional sine cardinal function, expressed as

$$\text{Peak}(i,j) = \left| \sin c\left(\frac{(i-i_0)^2}{2\sigma_i^2}\right) + \sin c\left(\frac{(j-j_0)^2}{2\sigma_j^2}\right)\right|, \tag{2}$$

where $\text{Peak}(i,j)$ defines the correlation peak value at (i,j) pixel in the output plane, $\sin c(x)$ is the sine cardinal function, and σ_i and σ_j denote the standard deviations for i and j around their respective mean values i_0 and j_0. The parameters in Eq. (2) were varied in order to get a large number of correlation planes (k) characterized with different correlation peak shapes $\text{Peak}_1, \text{Peak}_2, \ldots \text{Peak}_k$. Then, we expand these simulated correlation planes in terms of a set of orthonormal vectors. This is done by using the thin SVD matrix factorization to obtain the right singular vectors V_t [10]. The regressors Y^{peak} are given by

$$Y^{\text{peak}} = \begin{bmatrix} \text{thinSVD}(\text{Peak}_1) \\ \vdots \\ \text{thinSVD}(\text{Peak}_k) \end{bmatrix} = \begin{bmatrix} V_1^t \\ \vdots \\ V_k^t \end{bmatrix}. \tag{3}$$

The next step is to model the correlation plane noise. At first, we generate a face database for modeling the noise. Four subjects from the PHPID dataset were selected for this purpose, as shown in Fig. 3(a). Next, 52 facial images per person P(j), were obtained by rotating from right to left and from top to bottom. These images were correlated with a phase-only filter (POF: the chosen filter to validate the principle of our algorithm) formulated by choosing the face of the subject P(1) as shown in Fig. 3(a). Each of these correlations yields a noise realization noise_l ($1 \le l \le n$), where n denotes the number of noise models. In order to model the noise associated with subject P(1), the information signal (correlation peak) is suppressed while keeping the noise signal undisturbed. The full correlation plane is considered for the other faces. Using this analysis, we obtain the noise column vector, denoted as

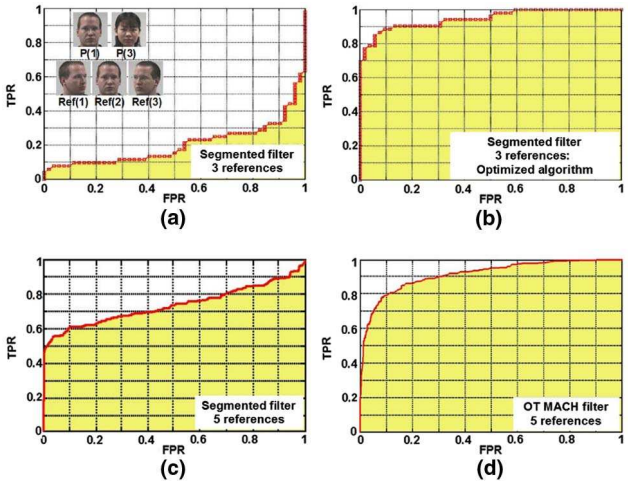


Fig. 3. (Color online) The results obtained using different kinds of filters. The first row shows the ROC curves using a 3-reference segmented filter: (a) without optimization, (b) with our optimized method and using 12 correlation plane noise models. The second row shows the corresponding results using 5-references filters: (c) using a segmented composite filter, (d) using a OT MACH filter.

$$\boldsymbol{Y}^{\text{noise}} = \begin{bmatrix} \text{noise}_1 \\ \vdots \\ \text{noise}_n \end{bmatrix}. \tag{4}$$

Using Eqs. (3) and (4), $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}^{\text{peak}} \\ \boldsymbol{Y}^{\text{noise}} \end{bmatrix}. \tag{5}$$

Now, the target correlation plane is decomposed thanks to the definition of the linear model of Eq. (1), without residual noise; i.e., weights corresponding to different types of regressors are calculated. The reconstructed correlation plane can be expressed as $\boldsymbol{P}'_c = \boldsymbol{Y}\boldsymbol{\beta}^+$, where P'_c represents the reconstructed correlation plane and β^+ represents the weights of the regressors. The residuals are obtained by calculating the difference between the target correlation plane P_c and the reconstructed correlation plane P'_c; i.e., $\boldsymbol{R} = \boldsymbol{P}_c - \boldsymbol{P}'_c$. Hence, the optimized correlation plane can be expressed by $\boldsymbol{P}_c^{\text{opt}} = \boldsymbol{Y}^{\text{peak}}\boldsymbol{\beta}^+_{\text{peak}} + \boldsymbol{R}$, where β^+_{peak} defines the weights of the regressors for the signal Y^{peak}.

The optimized correlation plane performance measure is the PCE. As will be shown shortly, one important advantage of this procedure is to lower the false alarm rate significantly. To illustrate the robustness of the proposed technique for face recognition, we tested it with a VanderLugt correlator using a 3-reference segmented filter [1] generated using the facial images of [1–3] as shown in Fig. 3(a).

Correlating subject P(1) with the segmented filter yields a true correlation, whereas correlating subject P (3) with this filter yields a false alarm. Fig. 3(a) shows the results obtained by correlating the 52 images of subject P(1) and subject P(3) with the 3-reference segmented filter. Fig. 2(b) represents the ROC curves obtained with 12 models for each subject with optimization. The reconstruction of the output correlation plane requires a processing time of 0.0883 s running Windows 7 (64-bit), 3.1 GHz CPU 4 GB RAM with the MATLAB software package. For comparison, Fig. 3(c) and Fig. 3(d) represent the results from the 5-reference filters using a segmented filter and an OT MACH filter, respectively [12]. From Fig. 3, we can see that without optimization, the TPR is close to 15% for a FPR set to 0%. Using the proposed technique, the VLC generates a TPR close to 70% of correct face identification with zero false positives. This recognition rate value exceeds the corresponding value for the optimized filter in [5]. To confirm the superior performance of our optimization, we compared it with an optimized 5-reference filter that yields a TPR close to 50% and 30%, respectively for the segmented and the trade-off maximum average correlation height (OT-MACH) filters. For both cases we found that the recognition rate was largely above 70%. As a benchmark for correlation plane quantification, it is instructive to compare the corresponding value of the area under curve (AUC) without optimization [Fig 3(a), which is equal to 0.201] and that obtained using our methodology [Fig. 3(b), which is 0.947] which is significantly larger.

Overall, our results dealing with different subjects indicate that the recognition rate tends to increase as the number of noise and peak models increases. To summarize, we have proposed a new LFM-SVD-based recognition algorithm for improved face recognition via correlation plane optimization. Test results obtained using this algorithm clearly indicate the effectiveness of this approach in relation to face recognition [4]. We have achieved recognition rates around 70% with 100% discrimination with the PHPID database. Our findings suggest that this algorithm can be adapted for different decision criteria and optimized correlation filters. Various tests show that it is possible to increase the robustness of the correlator while keeping a very good discrimination quality, typically comparable to that of the amplitude modulated phase only matched filter [3]. Furthermore, the simplicity of the proposed technique makes it is easier to implement. Future work may include investigation with other composite filters to show that this method is not segmented filter specific.

References

1. A. Alfalou and C. Brosseau, in *Face Recognition*, Milos Oravec, ed. (INTECH, 2010), pp. 354–380.
2. F. T. S. Yu and S. Jutamulia, *Optical Pattern Recognition* (Cambridge University, 1998).
3. A. A. S. Awwal, Appl. Opt. **49**, B40 (2010).
4. P. Katz, A. Alfalou, C. Brosseau, and M. S. Alam, "Correlation and independent component analysis based approaches for biometric recognition," in *Face Recognition: Methods, Applications and Technology*, (INTECH, to be published).
5. A. Alfalou and C. Brosseau, Opt. Lett. **36**, 645 (2011).
6. F. Dubois, Appl. Opt. **35**, 4589 (1996).
7. R. D. Juday, J. Opt. Soc. Am. A **18**, 1882 (2001).
8. H. Cardot, F. Ferraty, and P. Sarda, Statist. Probab. Lett. **45**, 11 (1999).
9. A. Reynaud, S. Takerkart, G. S. Masson, and F. Chavane, NeuroImage **54**, 1196 (2011).
10. M. Wall, A. Rechtsteiner, and L. Rocha, in *A Practical Approach to Microarray Data Analysis*, D. P. Berrar, W. Dubitzky, and M. Granzow, eds. (Springer, 2003), pp. 91–109.
11. N. Gourier, D. Hall, and J. L. Crowley, in *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures* (2004).
12. H. Zhou and T. H. Chao, Proc. SPIE **3715**, 394 (1999).

chapter

CORRELATION AND INDEPENDENT COMPONENT ANALYSIS BASED APPROACHES FOR BIOMETRIC RECOGNITION

P. Katz[1], A. Alfalou[1], C. Brosseau[2] and M. S. Alam[3]

[1] ISEN Brest, L@bISEN, 20 rue Cuirassé Bretagne CS 42807, 29228 Brest Cedex 2, France

[2] Université Européenne de Bretagne, Université de Brest, Lab-STICC, CS 93837, 6 avenue Le Gorgeu, 29238 Brest Cedex 3, France

[3] Department of Electrical and Computer Engineering, EEB 75, University of South Alabama,
6001 South Dr., Mobile, AL 36688-0002, USA.

ABSTRACT

Independent component analysis (ICA) models, describing a given signal as a linear combination of various independent sources, have proven to be a fruitful endeavor. One prominent example deals with audio applications in order to separate the speaker's voice from environmental noises disturbing it. However, very few ICA based systems are available for biometric encryption applications. For that specific purpose, the ICA method can be easily adapted to add noise to a target image in order to encrypt it. In this chapter, at first, we discuss biometric recognition systems based on the ICA and correlation approaches. Next, we explore an ICA-based algorithm for face recognition. Basically, it consists of building a base of independent components using a learning database that contains several chosen reference images. Then, the target image (image to be recognized) is projected on the independent component base, and the similarity between the target image and each of the reference images is studied. Discrimination tests between the proposed technique and alternate methods are conducted by using the Pointing Head Pose Image Database (PHPID). In this chapter we report some of the recent developments dealing with the ICA method for face recognition applications. As part of our analysis, we precisely determine a set of metrics aimed at better understanding the role of the number and choice of the reference images on the performance of the proposed technique.

INTRODUCTION

Over the last two decades tremendous advances has been made in face recognition techniques. This interest partly stems from potential applications in many diverse fields such as identification of wanted people in public areas, automation of passport registration at airports, and nonintrusive monitoring of activities of daily living, especially for the elderly [1].

Much progress has been made in the realm of face recognition techniques. A number of techniques can be employed based on optical correlation such as VanderLugt correlation (VLC), or on numerical methods such as eigenfaces [2], wavelets [3], principal component analysis (PCA) [4], and independent component analysis (ICA) [5]. Overall, ICA and PCA are versatile techniques allowing us to decompose any arbitrary image or signal into a linear combination of independent variables. The main difference between ICA and PCA arises from the fact that the former leads to independent images while the latter deals with uncorrelated images. This is an important issue in face recognition [6]. Assessing the performance of recognition techniques by relevant metrics is essential for understanding the advantages and limitations of these techniques.

In this chapter we report some of the recent developments dealing with the ICA method for face recognition applications. On the one hand, we consider a hybrid technique using ICA and correlation methods that was proposed recently by some of the authors [6]. On the other hand, a method based on ICA solely is investigated. Our main goal is to evaluate the receiver operating characteristic (ROC) [7] for characterizing the performances and limitations of this technique. Emphasis has been placed on the optical correlation methods using composite and segmented phase only filters (POFs) [8] and detection criteria of the correlation peak. In order to compare the performances of both methods, we report an extensive series of simulations aimed at better understanding the role of the number and choice of the reference images, and the pre-processing of the target images. Most importantly, the real power of the new method presented here is its high discrimination level for face recognition applications.

GENERALITIES

We start by discussing two standard approaches for face recognition approaches. The first method is based on VLC [9]. Single correlation, i.e. classical matched filter, POF, and multicorrelation, i.e. composite and segmented composite filters, approaches are considered [8]. Several correlation criteria are defined along with a performance metric of a classifier, i.e. the ROC [7]. Another hybrid method based on optical correlation and ICA will be introduced [8]. The method is designed to take advantage of the robustness of ICA and the discrimination of optical correlation.

VLC Technique for Face Recognition

VLC technique is based on the multiplication of the target image spectrum (image to be recognized) with a correlation filter H, fabricated with a reference image, as shown schematically in Figure 1.

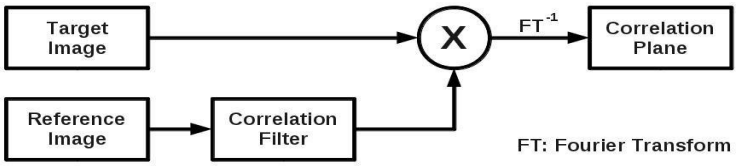


Figure. 1 Schematic diagram of the VLC method.

The scene containing the object to be recognized, called the input plane is multiplied by the correlation filter after applying the Fourier transform (FT) operation. Remarkably, the result obtained by applying the inverse Fourier transformation (FT-1) shows a central correlation peak (close to a two-dimensional sine cardinal), which is more or less narrow and intense depending on the resemblance between the target and reference images. The main advantage of this technique is that it can be easily implemented optically. Different filters have been designed for the purpose of increasing the discrimination and robustness of the VLC, originally based on the classical matched filter (CMF) [9]. For example, POFs have been largely explored by the image processing community since the early 1960s. Since, other methods were proposed to deal with multicorrelation [8]. That is to compare a target image with a set of reference images in order to be able to recognize a given subject for every position in space and face expression. The transfer function of the CMF takes the form

$$H_{CMF}(u,v) = \frac{\alpha S_{R_1}^{*}(u,v)}{N(u,v)}, \quad (1)$$

where $S_{R_1}^{*}$ corresponds to the reference image spectrum, N is the spectral density of the noise, and α is a constant. This filter is robust but cannot be used in practice since it is not discriminating.

Phase Only Filter

To provide a better discriminating filter, the POF was suggested [8]. The choice of this filter is based on the fact that the phase of a spectrum contains the necessary information to reconstruct the target image [10]. Hence, the spectrum amplitude which displays a fast dynamics but contains little information is unnecessary. To provide a better appreciation for this important property, an example of image is shown in Fig 2, with the original image (Figure 2a) and its phase (Figure 2b). The most distinct feature in Figure 2b is that it contains the information on the contours of the image. This allows us to get much more discriminating filters than the adapted filter.

Figure. 2 Illustrating phase selectivity of the POF. (a) Lena picture, (b) its phase representation.

The POF's transfer function reads as

$$H_{POF}(u,v) = \frac{S_{R_1}^*(u,v)}{\left|S_{R_1}(u,v)\right|}. \tag{2}$$

In contrast with CMF, POF leads to a very narrow correlation peak. This leads to a more discriminating filter with low robustness due to the strong sensitivity to image variations.

Composite Filter

In an attempt to overcome the shortcomings of CMF and POF, multicorrelation analysis was introduced. Its basic principle, shown in the diagram of Figure 3, is to fabricate a filter with several reference images, e.g. to deal with several face orientations. This permits to decrease the necessary number of correlations, and consequently the computation cost.

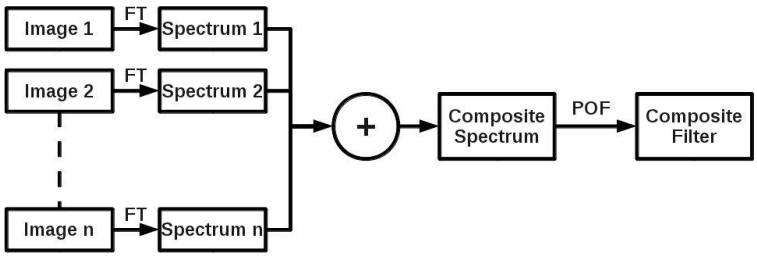


Figure. 3 Schematic diagram of the composite filter.

The composite filter is simply a linear combination of the n reference images. Its transfer function is

$$H_{COMP} = \sum_{i}^{n} a_i R_i, \tag{3}$$

where a_i is a weight coefficient. One advantage of this filter is that the correlation peaks of the reference images are additive, rendering it more robust to the rotation effects of the target image, allowing subject identification thanks to a larger face area analysis. However, it can lead to a saturation of the correlation plane when one considers too many reference images or when these images have high energy [11].

Segmented Composite Filter

The segmented composite filter, shown in Figure 4, can be used to deal with the correlation plane saturation of the composite filter. The segmentation of the Fourier plane is realized by assigning each pixel according a specific segmentation criterion, e.g. its energy, its spectrum gradient, its phase gradient, or its real part.

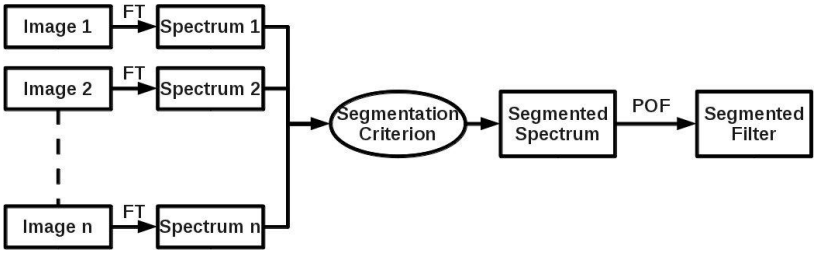


Figure. 4 Schematic diagram of the segmented composite filter.

The choice on the pixel is done according the inequality

$$a_i Crit^i_{(u,v)} \leq a_j Crit^j_{(u,v)} \; \forall \, j \in [\![1,n]\!] \; and \; j \neq i. \tag{4}$$

where $Crit^i_{(u,v)}$ corresponds to the segmentation criterion of the Fourier plane. For example, if one uses the image energy as segmentation criterion, the relative energy of each pixel after Fourier transforming each reference image is compared with the other images of the reference base. The resulting spectrum considers the high energy components of the base for each pixel. This filter, obtained after selection of the phase of the segmented image, allows us to use multicorrelation for a large set of reference images while avoiding the saturation of the Fourier plane.

Detection Criteria

Before turning to the presentation of face recognition methods, we discuss important prerequisite for a useful detection scheme.

Correlation Peak Detection Criteria

The comparison of the different algorithms used for implementing VLC [9] requires some useful ways to quantify and assess the correlation performances. We first offer a general discussion of the different metrics encountered in the literature for characterization of the correlation peak, e.g. signal-to-noise ratio (SNR), peak-to-correlation energy (PCE) [12], and their variants PCE' and PCE'' [13].

The SNR, and its counterpart expressed in dB, SNR_{dB}, are commonly employed metrics in signal processing. The simple definition of SNR is

$$SNR = \frac{Signal\ Power}{Noise\ Power} = \frac{C(Peak)}{\sqrt{\sum_{u=0}^{W}\sum_{v=0}^{H}|C_n(u,v)|^2}}, \tag{5}$$

where $C(Peak)$ is the maximal value of energy, (Peak denoting the coordinates of the correlation peak), $\sum_{u=0}^{W}\sum_{v=0}^{H} C_n(u,v) = \sum_{u=0}^{W}\sum_{v=0}^{H} C(u,v) - C(Peak)$ denotes the noise, i.e. the entire correlation plane $C(u,v)$ except the correlation peak, while W and H characterize the size of the Fourier plane. The SNR_{dB} reads as

$$SNR_{dB} = 10\log_{10}(SNR). \tag{6}$$

Concurrently, there are numerous efforts to develop other metrics to probe and understand the correlation performances. For example, PCE is defined as the energy of the

correlation peak normalized to the overall energy of the correlation plane [13]. The explicit definition follows a variant of the metric suggested by Dickey and Romero [14] :

$$PCE = \frac{\sum_{u=0}^{W} E_{Peak}\,(u,v)}{\sum_{u=0}^{W}\sum_{v=0}^{H} E_{Correlation\ \ Plane}\,(u,v)} = \frac{C(Peak)^2}{\sum_{u=0}^{W}\sum_{v=0}^{H} C(u,v)^2}, \tag{7}$$

where E_{Peak} and $E_{Correlation}$ denote respectively the energy of the correlation peak and of the correlation plane. Consequences of Eq. (7) are as follows. The energy of an intense correlation peak is much larger than the energy of the correlation plane. Hence, the value of the PCE will be large. In contrast, a wide correlation peak has a PCE value close to 0. It is worth observing that the correlation plane can have secondary peaks, which are not important in case of strong correlation, but which generate false alarms when the peak is not intense.

Two other convenient parameterizations to quantify and assess face recognition performances, i.e. PCE' and PCE'', were suggested by Horner [13]. The PCE' is chosen as a compromise between SNR and PCE

$$PCE' = \frac{C(Peak)}{\sum_{u=0}^{W}\sum_{v=0}^{H}|C_n(u,v)|^2}. \tag{8}$$

The PCE'' is defined as

$$PCE'' = \frac{PCE}{1-PCE}. \tag{9}$$

Receiver Operating Characteristic

We next focus on the ROC representation. In practical calculations, face recognition can be modeled as a two-class prediction problem (binary classifier system). Either, the subject is recognized as that chosen for the fabrication of the filter, or it is not recognized. For our purpose, we define the vector $\underline{w}$, also termed the observation vector, composed of n observations $w_1, \cdots, w_n$. The basic idea is to make the best decision from a set of observations according to a criterion $\hat{\theta}(\underline{w})$ which is the best estimate of the parameter θ. We first define $\underline{Y}$ as the vector formed by the ensemble of values taken by the evaluation criterion when the target image corresponds to the reference subject, and $\underline{Z}$ the vector formed in the opposite case. From basic discrete random variable (pi, yi) theory it follows that the expected value of $\underline{Y}$ is given by

$$E[\underline{Y}] = \mu = \sum_{i=1}^{n} p_i y_i = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{10}$$

and the standard deviation is $\sigma = \sqrt{E[\underline{Y}^2] - E[\underline{Y}]^2}$. When $\underline{Y}$ and $\underline{Z}$ are Gaussian distributed, one posits that $\underline{Y} \sim \mathcal{N}(\mu_{\underline{Y}}, \sigma_{\underline{Y}}^2)$ and $\underline{Z} \sim \mathcal{N}(\mu_{\underline{Z}}, \sigma_{\underline{Z}}^2)$ (Figure 5). Four cases can occur: detection, non-detection false alarm, and false non-detection. To evaluate graphically the performance of the classifier, the remaining task is now to determine the ROC curve [7], which is a graphical plot of the true positive rate TPR (called also sensitivity, or probability of true detection) versus false positive rate FPR (called also 1-specificity, or false alarm probability) as its discrimination (detection) threshold is varied. Lets us call H_0 the hypothesis such that the target image is that of subject 0, H_1 is that of subject 1, D_0 that subject 0 is detected, and D_1 that subject 1 is detected. We have

$$FPR = P(D_1|H_0) + P(D_0|H_1) \text{ and } TPR = P(D_0|H_0) + P(D_1|H_1). \tag{11}$$

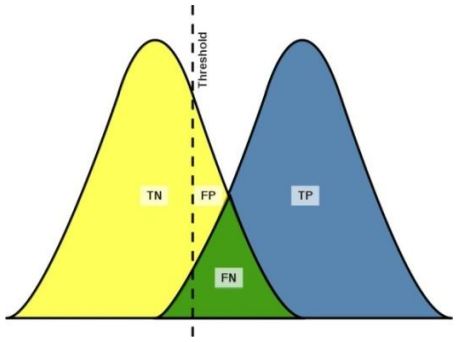


Figure. 5 Gaussian curves. TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

Now if we consider that a decision is made by comparing the estimator $\hat{\theta}(\underline{w})$ with a given threshold s [7], i.e.

$$\hat{\theta}(\underline{w}) \underset{H_{1-i}}{\overset{H_i}{\gtrless}} s, i \in \{0,1\}. \tag{12}$$

we immediately obtain FPR and TPR as

$$FPR = P\left(\hat{\theta}(\underline{w}) > s \middle| H_1\right) + P(\hat{\theta}(\underline{w}) < s|H_0)$$

and

$$TPR = P\left(\hat{\theta}(\underline{w}) > s \middle| H_0\right) + P\left(\hat{\theta}(\underline{w}) < s \middle| H_1\right). \tag{13}$$

In practice, s is varied between 0 and 1. According the values of the estimation $\hat{\theta}(\underline{w})$ and threshold s, the classifier makes a decision, D_0 or D_1, leading to a 2×2 confusion matrix, or contingency table (Table 1).

Table. 1: Confusion matrix or contingency table.

		Positive	Negative
Positive		TP	FP
Negative		FN	TN

Finally, we define the false positive (FPR) and true positive (TPR) rates as

$$FPR = \frac{FP}{FP + TP} \text{ and } TPR = \frac{TP}{TP + FN}. \tag{14}$$

The ROC curve is obtained by plotting the FPR versus the TPR for each threshold value s (Figure 6). The perfect classification, corresponding to the case for which the densities of probability of $\underline{X}$ and $\underline{Y}$ have disjoint supports, would yield a point in the upper left corner of the ROC space, i.e. representing 100% sensitivity (no false negatives) and 100% specificity

(no false positives). An important property of this representation is that the diagonal line corresponds to random guess. This diagram discriminates between a region of good classification (points above the diagonal) and a region of poor classification (points below the diagonal). An alternative way for representing the performance of a classifier is to consider the area under curve (AUC) [15]. In the case of a random classification $AUC = 5$ whereas $AUC > 5$ for a good classification.

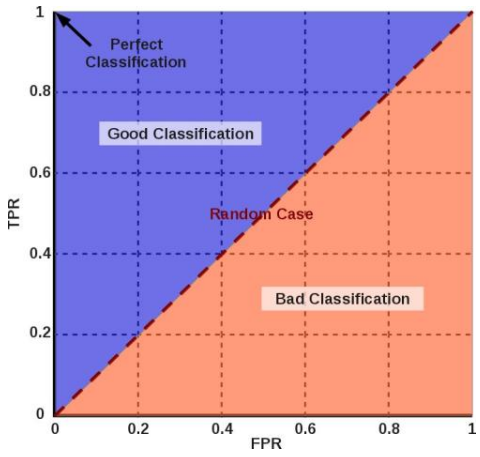


Figure. 6 The ROC space.

Robust and Discriminating Method for Face Recognition Based on Correlation Technique and ICA Model

It is well established that optical correlation techniques are very sensitive to face rotation. The segmented composite filter was proposed to deal with this problem but its performance decrease as the number of reference images is increased [11]. To overcome this problem, several computational strategies have been suggested, including ICA [5]. Other fields of research where such ICA-based approaches have been fruitfully applied are acoustics [16], cognitive sciences [17-19], optical encryption [20], and recognition [21-22].

Principle of ICA

In its simplest form, the ICA method predicts, from a set of n observations, $S_1, \cdots, S_n$, statistically independent components $C_1, \cdots, C_n$, such the observations can be represented by a linear combination of the different components. Hence, for the vectors $\underline{S}$ and $\underline{C}$, of dimension n, and A being the coefficient matrix of dimension n^2, the decomposition of $\underline{S}$ reads as

$$\begin{cases} s_1 = a_{1,1}c_1 + a_{1,2}c_2 + \cdots + a_{1,n}c_n \\ s_2 = a_{2,1}c_1 + a_{2,2}c_2 + \cdots + a_{2,n}c_n \\ \vdots \\ s_n = a_{n,1}c_1 + a_{n,2}c_2 + \cdots + a_{n,n}c_n \end{cases}, \tag{15}$$

with

$$\underline{S} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}, \underline{C} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}, \mathrm{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix},$$

or, can be expressed in the entirely equivalent matrix form

$$\underline{S} = A\underline{C}. \tag{16}$$

The basic objective of this method is to find the linear combination which minimizes the statistical dependence between its components. For that specific purpose, one is searching from the knowledge of $\underline{S}$ the matrix W such $W \times A$ is diagonal. These components are given by

$$\underline{C} = W\underline{S}. \tag{17}$$

Several algorithms, adapted for ICA, were proposed in the literature, e.g. fastICA [23], using criteria such as kurtosis, negentropy, and minimization of mutual information.

ICA and Correlation Based Method

We now turn our attention to the numerical procedure involving both ICA and VLC [6]. Most interestingly, in this analysis the system still retains the robustness of ICA and the high discrimination of VLCs. The basic principle of this analysis is to recognize a target image by using a set of reference images. The ICA is used as pre-processing stage of the reference base. Image recognition is realized thanks to a simple 1-reference POF. It is worth noting that this method can be implemented optically or numerically.

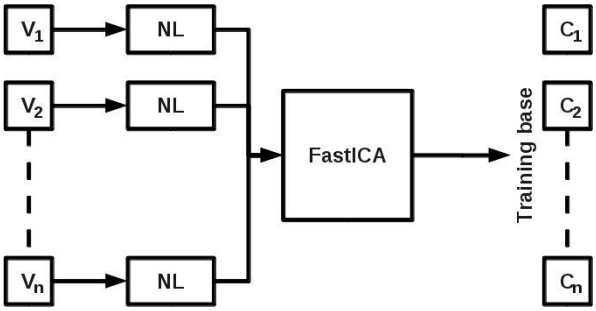


Figure. 7 Definition of the independent component learning base.

Our algorithm can be summarized as follows. The first step [6], shown graphically in Figure 7, consists in fabricating a reference base from a learning base of n different images $(V_1, \cdots, V_n)$ of subject X. Here, we used the fastICA algorithm [23]. The use of a specific nonlinear function over the learning base is also necessary to ensure that the algorithm is convergent [20]. The next step consists to correlate each image V_i of the learning base to the different POFs, fabricated from the base of independent components C_j. From these correlation planes, we get

$$PCE_{ij} = f\left(V_i \otimes pof_j\right), \tag{18}$$

where f denotes is a specific function, and pof_j is the inverse FT of the POF fabricated from C_j (Figure 8).

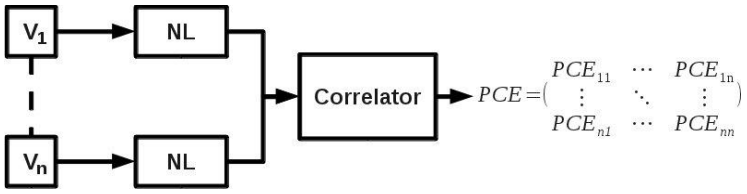


Figure. 8 Definition of the PCE matrix.

The last step of the algorithm, displayed in Figure 9 consists of the comparison and recognition of the target image. We consider a target face $V_?$ of a given subject. Next, we decompose it in a linear combination of independent components of the reference base. In a first step, this target image is correlated with every pof_j of the base. Then, we get a vector called Val_{Corr}, such that $Val_{Corr} = (PCE_1^{'}, \cdots, PCE_n^{'})$, with $PCE_j^{'} = f(V_? \otimes pof_j)$. Finally, we define the components of the as

$$\varepsilon_i = |PCE_1^{'} - PCE_{i1}| + |PCE_2^{'} - PCE_{i2}| + \cdots + |PCE_n^{'} - PCE_{in}|. \quad (19)$$

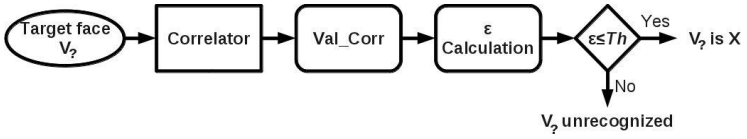


Figure. 9 Illustration of the recognition procedure.

Finally, the arithmetic mean ε of the components of the error vector is compared to a given threshold Th. The target image $V_?$ is recognized as arising from subject X when ε is smaller than the threshold and is not recognized otherwise. This technique offers many advantages. Firstly, the architecture is both optically and numerically implementable.

Secondly, it leads to better results than those obtained with a simple composite filter [6], using a simple 1-reference POF, i.e. a single matrix of independent components. Thus, outstanding performances are expected for a multicorrelation implementation.

ICA BASED BIOMETRIC RECOGNITION

Previously, we have introduced a method allowing us to fabricate an image base of statistically independent references and applied it to an optical correlation technique such as VLC in order to improve the face recognition performances. Now, we consider a purely numerical technique based on ICA which is found to be robust for biometric recognition such as face recognition [5].

Method

Principle

Below we describe in detail the 4-step algorithm which is central to our analysis. Firstly, a learning base of reference images is fabricated. Secondly, the ICA method is applied over this base leading a coefficient matrix A and a matrix of independent components C: each line of A corresponds to a specific reference image. Thirdly, $V_?$ is written as a linear combination of the $C_j's$. Fourthly, the coefficients encoded in $\underline{A'}$ are compared with those obtained from the decomposition of the target image with each line of the matrix A, leading to the error between the target image $V_?$ and each reference image. Finally, the smallest error corresponds to the most resembling reference image to the target image. That is, if the error is less than a threshold, then the image corresponds to the reference subject. The central point of our analysis, shown in Figure 10, is to decompose a base $\underline{V}$ of reference images of a subject X with the fastICA method [23]. Then, we get a vector $\underline{C}$ of independent components and a matrix A of coefficients (Figure 10).

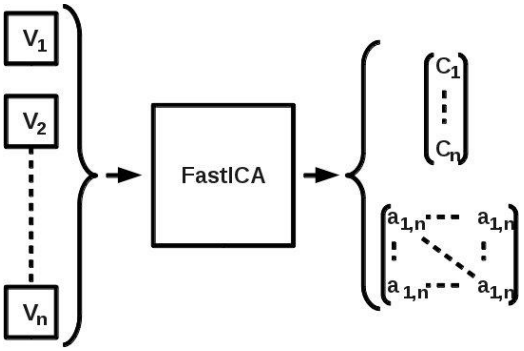


Figure. 10 Principle of decomposition.

The second step of the algorithm, with reference to Figure 11, considers a target image $V_?$ corresponding to an unknown subject. This image will be written as a linear combination of independent components forming $\underline{C}$ calculated within the learning base. Then, $V_?$ is written in the form

$$V_? = \underline{A'C}, \tag{20}$$

where $\underline{A'}$ of size n, i.e. $\underline{A'} = (a_1', \cdots, a_n')$, summarizes the set of coefficients allowing us to reconstruct $V_?$ from the components of $\underline{C}$. In this way, we find $\underline{A'} = V_? C^{-1}$. In order to a make decision, the differences between the vector $\underline{A'}$ and each line of the matrix A are calculated (Eq. 21). Minimizing the error function leads to the most resembling image of the reference base to the target image $V_?$.

$$\begin{vmatrix} \varepsilon_1 & = & |a_1' - a_{1,1}| + |a_2' - a_{1,2}| + \cdots + |a_n' - a_{1,n}| \\ \vdots & & \vdots \\ \varepsilon_n & = & |a_n' - a_{n,1}| + |a_2' - a_{n,2}| + \cdots + |a_n' - a_{n,n}| \end{vmatrix}. \tag{21}$$

In a similar fashion as was mentioned above the error is defined as $\varepsilon = (\varepsilon_i + \cdots \varepsilon_n)/n$. When the error is less than a given threshold the target image $V_?$ is considered as representing subject X, otherwise it is not recognized. On the one hand, this algorithm relies on the strong

robustness of the ICA method. On the other hand, it allows a precise quantification of the resemblance of the target image with the reference base.

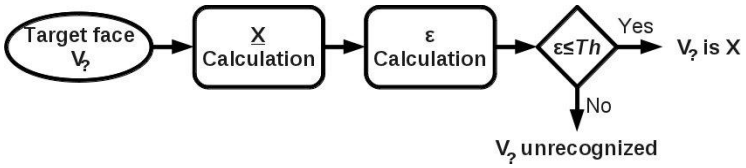


Figure. 11 Recognition procedure.

Protocol

Next, we validate this approach by a series of tests using subjects 1 and 2 of the PHPID [24]. This base has 15 subjects, 93 references per subject, and the angle characterizing the face orientation varies every 10°, either horizontally or vertically. Here, we shall consider only 53 images per subject, i.e. the first 20 and last 20 images have been removed. In addition, the face images were reframed (215×215 pixels), to minimize the background noise of the images (Figure 12). The images of subject 1 were chosen for fabricating the reference base (Figure 12 (a)). The recognition procedure is tested for both subjects 1 and 2 (Figure 12 (a-b)).

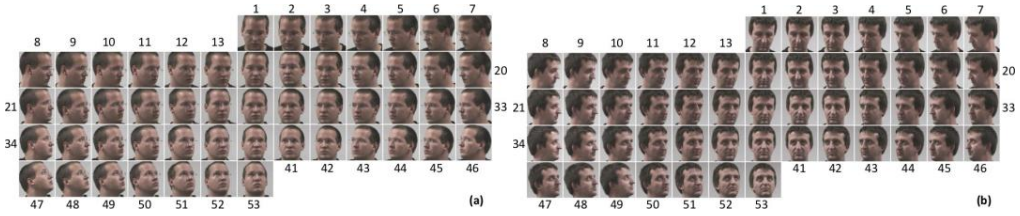


Figure. 12 Database 1 : (a) Subject 1 and (b) Subject 2.

As concerns the study of the noise of the input image, 8 target images of the PHPID will be considered (Figure 13). Here, the images were not reframed. Two kinds of test were performed. In the first test, subject 1 was the reference subject (Figure 13 (a-f)). For the second we test used two images (Figure 13 (g-h)), i.e. the two faces are similar to the reference face (Figure 13 (g)), or one face is different than the reference one (Figure 13 (h)).

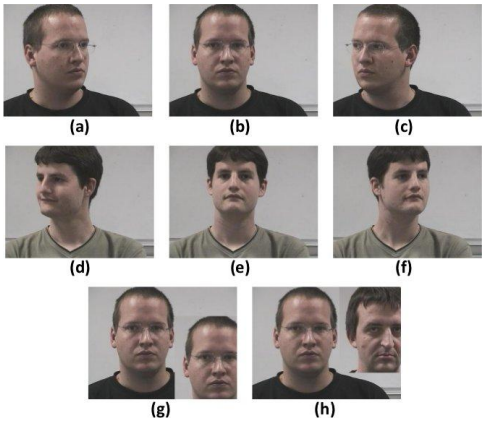


Figure. 13 Database 2: (a-c) Subject 1, (d-f) Subject 2, (g-h) Mixed images.

Experimental Results

Influence of the Reference Images

The influence of the choice of the reference images is illustrated in Figure 15. Four ROC curves are shown and were obtained from using different sets of reference images of Figure 14, i.e. images 1, 27, and 53 for Figure 15 (a-b), images 26, 27, and 28 for Figure 15(c-d), images 27, 30, and 36 for Figure 15 (e-f), and images 9, 27, and 45 for Figure 15 (g-h).

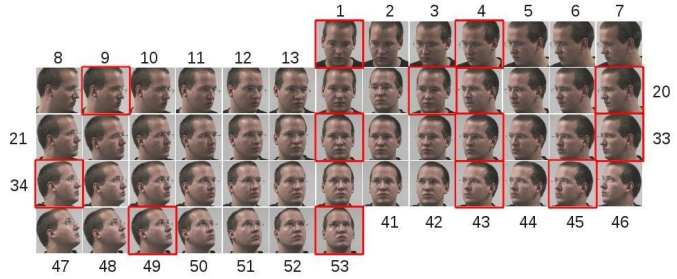


Figure. 14 Subject 1. The images framed in red are the reference images.

We can immediately observe that some image configurations lead to poor results. For example, choosing the images 1, 27, and 53 as reference images gives a ROC curve (Figure 15 (a)) going well below the random guess line, with a true recognition rate of 30% for a false alarm rate of 60%. If the FPR is set to 10% the true recognition rate is only of 19%. Additionally, choosing images 27, 26, and 28 gives a ROC curve (Figure 15(c)) close to the random guess line and a TPR of 19% when the FPR is set to 10%. The situation is quite different if the images 27, 30, and 36 (Figure 15(e)); or the images 9, 27, and 46 (Figure 15(h)) are selected, the ROC curve is now above the random guess line. The former case leads to a TPR of 52% when the FPR is set to 10%, while the latter case gives a TPR of 40% for a FPR set to 10%. These numerical results prove that the choice of the reference images has a strong influence on the performances of this ICA-based algorithm.

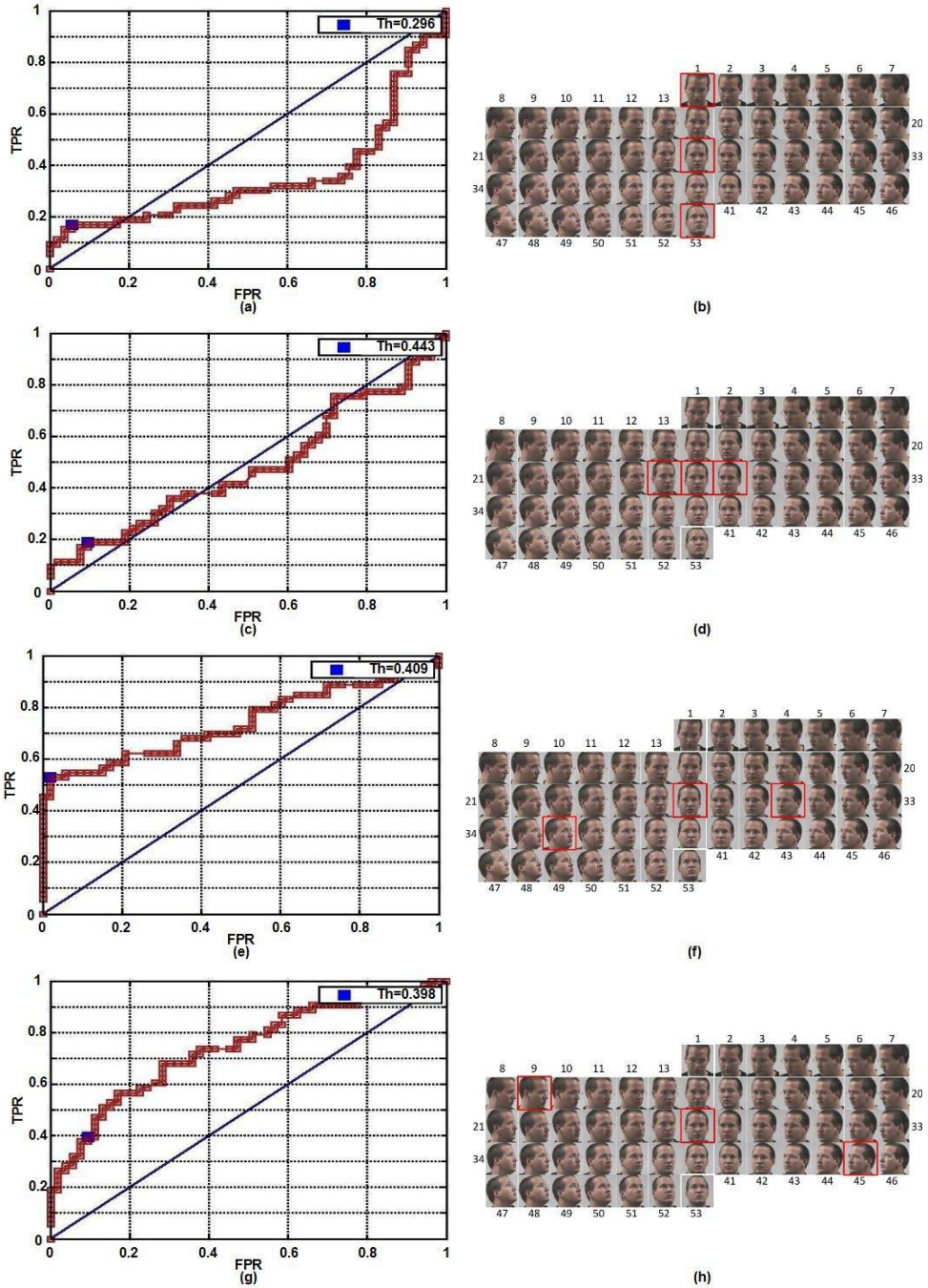


Figure. 15 Simulation results. (a-b) ROC plot and the three references chosen (references 1, 27, and 53). (c-d) The same as in (a-b) with references 26, 27, and 28. (e-f) The same as in (a-b) with references 27, 30, and 36. (g-h) The same as in (a-b) with references 9, 27, and 45.

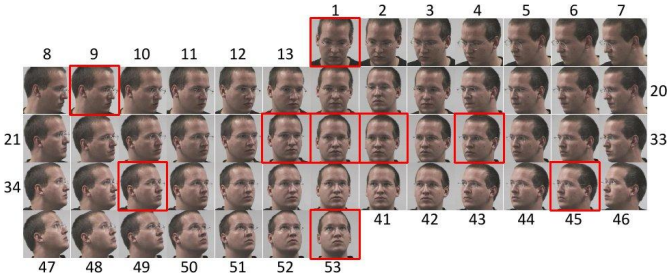


Figure. 16 Subject 1. The images in red are the references.

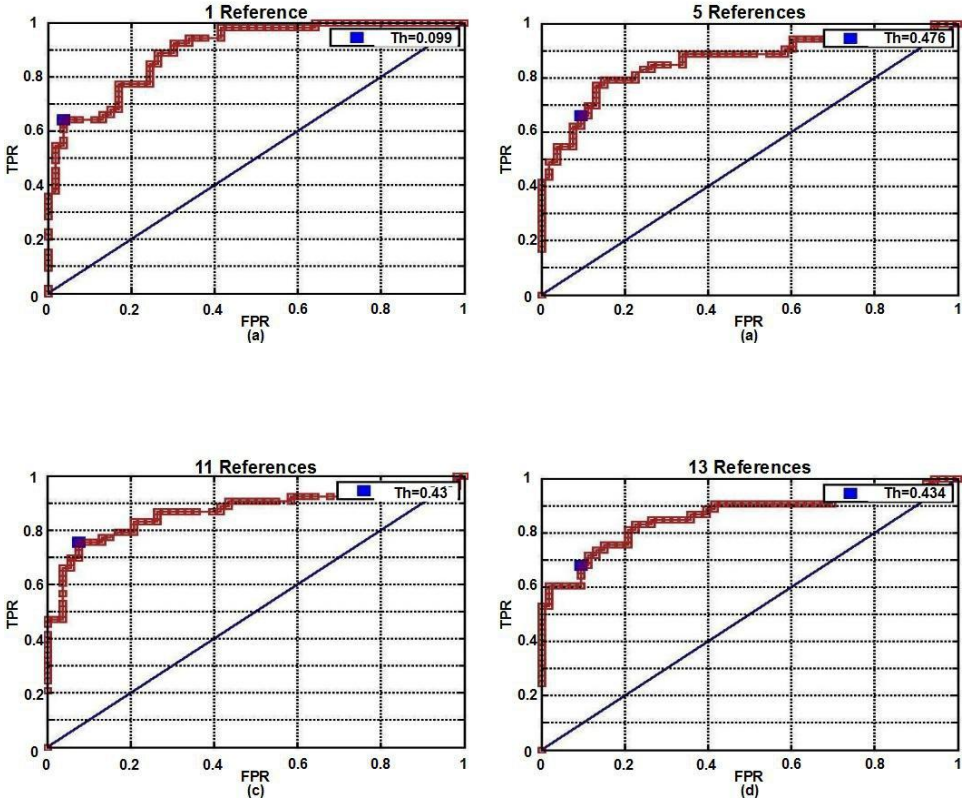


Figure. 17 Simulation results. (a) ROC plot using 1 reference. The blue square represents the best TPR value for a FPR set to 10%. (b), (c) and (d) are the same as in (a) for respectively 5, 11 and 13 references.

Influence of the Number of Reference Images

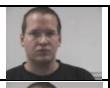
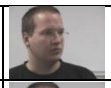
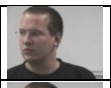
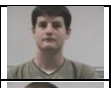
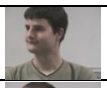
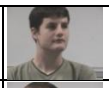
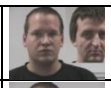
To further characterize the influence of the number of reference images we shall use 1, 5, 11, and 13 reference images among those of subject 1 (Figure 16). Figure 17 summarizes the behavior of the ROC curves corresponding to these assumptions.

From visual inspection, the classification is less efficient as the number of references is increased, i.e. we get a TPR of 64% for a FPR set to 10% and a single reference and a TPR of 51% for a FPR set to 10% and 5 references. This may be due to the specific choice of reference images, or to the saturation of the filter (the image 27 is used as reference for both filters). If the number of reference images is increased, the TPR increases to respectively 75% and 68% for 11 and 13 reference images and FPR set to 10%. Other keys to this work are the fabrication and recognition times of the reference base which depend of the number of images used. For a base containing a single reference the fabrication time is found to be $2.7\ s$ and the recognition time is $0.43\ s$. When 13 reference images are used these times they are respectively equal to $5.3\ s$ and $0.6\ s$. We interpret this behavior to be due to the size of the matrix of independent components which increases as the number of images is increased.

Effect of Noise in the Target Image

The effect of the background of the target images is first investigated. The images shown in Figure 13 are used as target images of the algorithm and the first 3 images (Figure 13 a-c) are taken as references. The results of our investigation are listed in Table 2. The first line of this table corresponds to the target images, the second line shows the reference images with minimal error (RIME), and the third line contains the error. We observe that the ICA-based method is strongly robust for distinguishing the two subjects (when the target image has only one face). The error is found of the order of 10^{-31} when the target and reference images are identical and of 10^{-6} otherwise. Although both errors are very weak the 25 orders of magnitude difference between these numbers lead to an efficient recognition technique. When a second face is introduced in the target image the error is found to increase significantly up to 10^{-6} even if both faces correspond to the same reference subject. This is an example of the low robustness to noise of the ICA-based method.

Table 2: ICA errors using target images shown in the first line. RIME means reference image with minimal error.

Target image								
RIME								
Error	$1.3 \cdot 10^{-31}$	$1.3 \cdot 10^{-31}$	$2.7 \cdot 10^{-33}$	$7 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$7.3 \cdot 10^{-6}$	$2.7 \cdot 10^{-6}$	$1.1 \cdot 10^{-6}$

COMPARISON BETWEEN ICA MODEL AND VLC USING AN OPTIMIZED COMPOSITE FILTER FOR FACE RECOGNITION

We now describe the recognition performances of the ICA and VLC methods. Our first aim is to examine different filters used to implement the VCL. Our second aim is to present the results. The two approaches based on ICA and VLC will be compared and their performances discussed below.

Filter Fabrication

Phase Only Filter

Following the definition of the POF shown in section 2, the algorithm was tested with subjects 1 and 2 of the PHPID [24] (Figure 12). In order to compare with the ICA-based algorithm, the effect of the noise of the input image will be also studied as above (Figure 13). The image of the central position of subject 1 (Figure 18) is used as reference image for the fabrication of the filter.

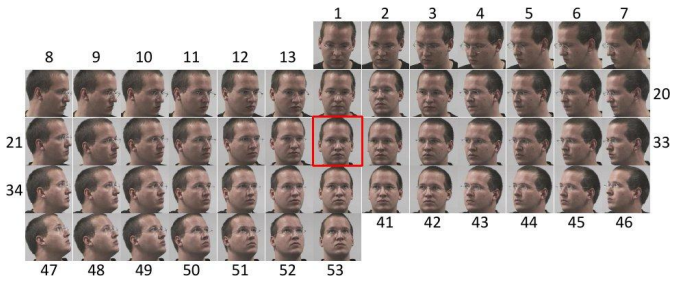


Figure. 18 Subject 1. The reference image is framed in red.

Segmented Composite Filter

As described previously, the segmented composite filter is a filter composed of many reference images. The principle of this method is to choose, for all spectra of the reference images, pixels containing most of the information for the given application. Each pixel of the reference base will be compared to the others for each image, in order to select the one corresponding to the segmentation criterion. For a given position (u, v) in the Fourier plane, the energy of the pixel is compared for the entire set of reference images, and is given by $E_{(u,v)} = I^2_{(u,v)}$. $I_{(u,v)}$ is the amplitude of the pixel and $E_{(u,v)}$ is its energy at coordinates (u, v). Note that the energy of each pixel is normalized by the overall energy of the image in the spectral domain. Hence, the spectrum of the segmented composite filter is composed of the most energetic pixels of the entire set of reference mages. The energy criterion reads

$$a_i \frac{E^i_{(u,v)}}{\sum_{u=0}^{W}\sum_{v=0}^{H} E^i_{(u,v)}} \geq a_j \frac{E^j_{(u,v)}}{\sum_{u=0}^{W}\sum_{v=0}^{H} E^j_{(u,v)}}, \forall\, j \in [\![1, n]\!] \text{ and } j \neq i, \tag{22}$$

where $E^i_{(u,v)}$ denotes the energy at coordinates (u, v) of the ith image and $\sum_{u=0}^{W}\sum_{v=0}^{H} E^i_{(u,v)}$ represents the energy of the ith image. The major drawback of this criterion is that it does not

take into account of the phase [10]. In order to consider the phase information one can use the complex gradient of the spectrum, given by the partial derivatives of the spectrum with respect to coordinates u and v

$$\nabla \mathrm{S}_{(\mathrm{u,v})} = \begin{pmatrix} \dfrac{\partial \mathrm{S}_{(\mathrm{u,v})}}{\partial u} \\ \dfrac{\partial \mathrm{S}_{(\mathrm{u,v})}}{\partial \mathrm{v}} \end{pmatrix} \tag{23}$$

where $S_{(u,v)} = A_{(u,v)}e\hat{}(i\varphi)$ denotes the spectrum at coordinates (u, v). The square of gradient modulus,

$$\left|\nabla S_{(u,v)}\right|^2 = \left|\frac{\partial S_{(u,v)}}{\partial u}\right|^2 + \left|\frac{\partial S_{(u,v)}}{\partial v}\right|^2, \tag{24}$$

strongly depends on the phase. The segmentation is now realized according the criterion

$$a_i\left|\nabla S^i_{(u,v)}\right| \geq a_j\left|\nabla S^j_{(u,v)}\right|, \forall\, j \in [\![1, n]\!] \text{ and } j \neq i. \tag{25}$$

For each pixel of the overall reference base, the one having the largest magnitude of the gradient will be selected for fabricating the segmented composite filter. In contrast with the energy criterion, the phase and the modulus of the spectrum are taken into account for fabricating the filter. It is also instructive to investigate the phase gradient defined as the phase derivative with respect to the pixel

$$\nabla \varphi_{(\mathrm{u,v})} = \begin{pmatrix} \dfrac{\partial \varphi_{(\mathrm{u,v})}}{\partial u} \\ \dfrac{\partial \varphi_{(\mathrm{u,v})}}{\partial \mathrm{v}} \end{pmatrix}, \tag{26}$$

where $\varphi_{(u,v)}$ denotes phase at coordinates (u, v). The segmentation is now realized according the gradient modulus, leading to

$$a_i\left|\nabla \varphi^i_{(u,v)}\right| \geq a_j\left|\nabla \varphi^j_{(u,v)}\right|, \forall\, j \in [\![1, n]\!] \text{ and } j \neq i. \tag{27}$$

In contrast with the energy criterion, the gradient criterion does not consider the modulus of the spectrum, and thus has a slow dynamics. Another segmentation criterion considers the real part of the pixel. The pixel selection is done according the condition

$$a_i \frac{A^i_{(u,v)} \cos\left(\varphi^i_{u,v}\right)^2}{\sum_{u=0}^{W}\sum_{v=0}^{H} E^i_{(u,v)}} \geq a_j \frac{A^j_{(u,v)} \cos\left(\varphi^j_{u,v}\right)^2}{\sum_{u=0}^{W}\sum_{v=0}^{H} E^j_{(u,v)}}, \forall\, j \in [\![1, n]\!] \; and \; j \neq i. \tag{28}$$

Note that, in a similar way as is done for the energy criterion, a normalization of the pixel by the overall energy of the image is realized. The main advantage of this criterion is to consider both phase and modulus.

Experimental Results

Our simulation scheme of the VLC is now used to explore the performances of the POF for the target image base composed of subjects 1 and 2 (Figure 12).

Phase Only Filter

In order to assess the POF performances, two filters fabricated with the central images of the reference subject for the two bases 1 and 2 (Figure 19) have been applied to all target

images shown in Figure 12. The results are shown in Figure 19. The corresponding ROC curve (Figure 19 (b)) indicates a TPR value of 20% for a FPR set to 0%. When the FPR is set to 9 %, the TPR is equal to 58 %. These low values are consistent with the PCE values shown in Figure 19 (a), i.e. the value of the PCE for the reference image, corresponding to a high correlation, is very large ($4.4 \cdot 10^{-3}$) compared to the values of the PCE for the other images of subject 1 ($2 \cdot 10^{-4}$). Consequently, the values of the PCE are very similar for most of the images of subject 1, even smaller than the values of the PCE for the images of subject 2, thus leading to classification mistakes. POF is very sensitive to small changes in face rotation: it has a high discrimination power but has a low robustness.

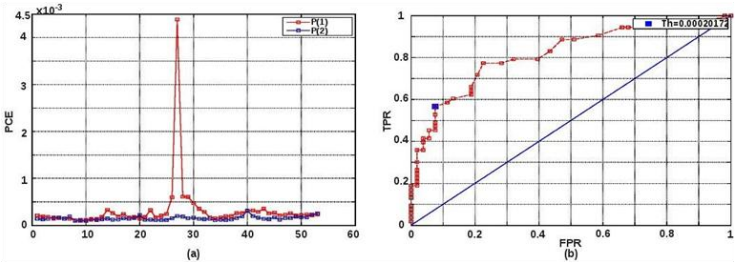


Figure. 19 Results using a POF, (a) PCA values, (b) ROC curve.

Next, we have investigated the effect of background noise of the target and reference images. For that purpose, the images shown in Figure 13 have been used for characterizing the performance of the POF. The first image (Figure 13 (a)) is taken as reference. The data are presented in Table 3. The first line corresponds to the target images, the second and third lines contain, respectively, the correlation planes and the values of the PCE. For the first 3 target images of Table 3 (identical reference and target images), it is to be noted that the correlation planes have an intense correlation peak with low level of noise. In contrast, when the target and reference subjects are different the correlation peak broadens and is characterized by a high level of noise. Although the value of the PCE is smaller for the second target face than for the first one, a correlation peak is observed when the target and reference subjects are different. The values of the PCE are close whether correlation is effective or not. As a result, this causes the increase of the false alarm rate. By contrast, the results obtained with the images containing background noise are interesting. When the target images contain a second face (the last two images of Table 3) we find that the correlation planes have an intense peak with low level of noise and PCE values which are comparable with those of the first 3 images of Table 3. To summarize, the background noise of the target image has little effect on the correlation result. The POF is strongly robust to the noise related to the background of the image.

A POF fabricated with the central image of subject 1 (Figure 18) has been applied to the base of reference images (Figure 12). The correlation peak was characterized by the different criteria. The results, shown in Figure 20, are very similar. In addition, the ROC curves shown in Figure 20 show the same trend. That is, for a FPR set to 0%, the TPR is close to 20% for every criterion. It should be noted that only small differences are noticeable in Table 4. Note that the SNR is less performing than the other criteria. Combined to its robustness to noise, the PCE' is the most efficient criterion [13].

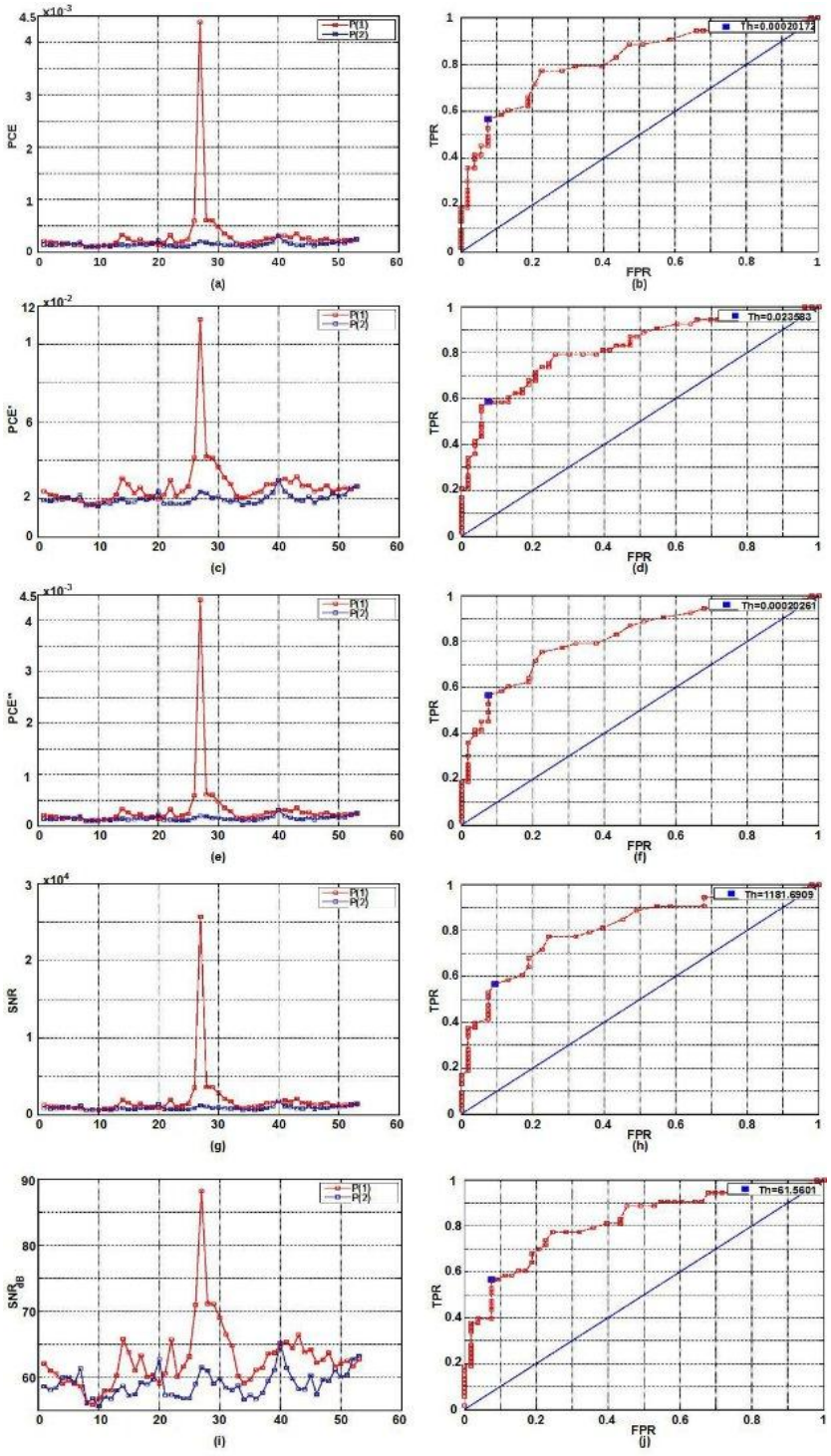


Figure. 20 Simulation results using a POF and different metrics. (a) PCE, (c) PCE', (e) PCE'', (g) SNR, (i) SNR_{dB}. The corresponding ROC curves are shown in (b), (d), (f), (h) and (j).

Table 3: Correlation planes using target images shown in the first line.

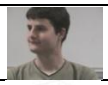
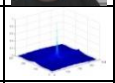
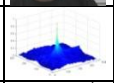
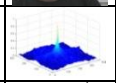
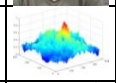
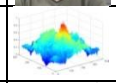
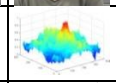
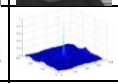
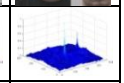
Target image								
POF								
PCE	$3.4 \cdot 10^{-3}$	$9.1 \cdot 10^{-4}$	$9 \cdot 10^{-4}$	$6.9 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$	$3.7 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$

Table 4: TPR and FPR values for the different peak detection criteria.

Criterion	Threshold	TPR (%)	FPR (%)
PCE	$2 \cdot 10^{-4}$	56.6	7.6
PCE'	$2.4 \cdot 10^{-4}$	58.5	7.6
PCE''	$2 \cdot 10^{-4}$	56.6	7.6
SNR	$1.2 \cdot 10^{-3}$	56.6	9.4
SNR_{dB}	62	56.6	7.6

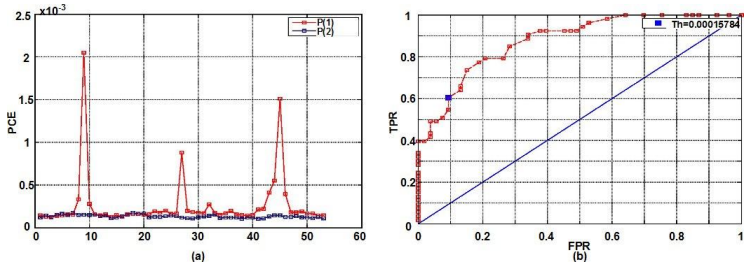


Figure. 21: PCE results and ROC curve using a segmented composite filter.

Segmented Composite Filter

The segmented composite filter, fabricated with images 9, 27, and 43 of the reference subject (Figure 16) was applied to the set of target images shown in Figure 12. The segmentation criterion chosen was the energy of the spectrum. As was shown in Figure 19 (a), the values of the PCE for the set of target images are weak. We find that several values of the PCE corresponding to the target images of the reference subject are less than the values of the PCE of the target images of the second subject. This has for effect to decrease the performances of this filter. Consequently, one may introduce a couple of supplementary reference images to increase the robustness of this filter. Figure 21 shows the results obtained with a segmented composite filter using 3 reference images. The recognition performances are well compared to those of the POF since the TPR is now equal to 40% when the FPR is set to 0%. This is due to the large increase of the PCE corresponding to the added reference images as is evidenced in Figure 21 (a). Increasing the number of reference images has for effect to enhance the performances of the classifier. In addition, the recognition time remains identical since only one correlation is needed, even if the fabrication time of the filter is a little bit larger.

From a trial and error method we noted that 13 reference images were sufficient to recognize a large part of the base. Several implementation of this 13-reference correlation were tested, e.g. a single filter segmented with all 13 references, or several filters containing some of these 13 references. The filters were fabricated using the images shown in Figure 16. They were segmented with the energy criterion and a weighting factor set to 4 for the central image. Several tests were made with 6 filters with 3 references, 3 filters with 5 references, 2 filters with 7 references, and 1 filter with 13 references (the image 27 having the central position is systematically used for fabricating the filters). The results shown in Figure 22 lead to better performances than those obtained previously (Figure 21) and using 3 reference images, i.e. a minimum value of 62% is obtained for the TPR when the FPR is set to 0% (Figure 22 (d)). In addition, the minimum value of the TPR is 79% when the FPR is set to 10%, to be compared with the 60% obtained earlier. The effect of the number of filters on the recognition performance is noticeable. For example, if the FPR is set to 0% the TPR is respectively of 79%, 62%, 83%, and 77% for 6, 3, 2, and 1 filter. If the FPR is set to 10% these numbers are 89%, 79%, 89%, and 83%. Using 6 filters with 3 references and 2 filters with 7 references lead to the best results. The weak performances of the filter with 13 references are due to a saturation effect. Using a combination of 2 filters with 7 references requires 2 correlations while using a combination of 6 filters necessitates 6 correlations. Hence, it is the best compromise between accuracy in the recognition and computational expense.

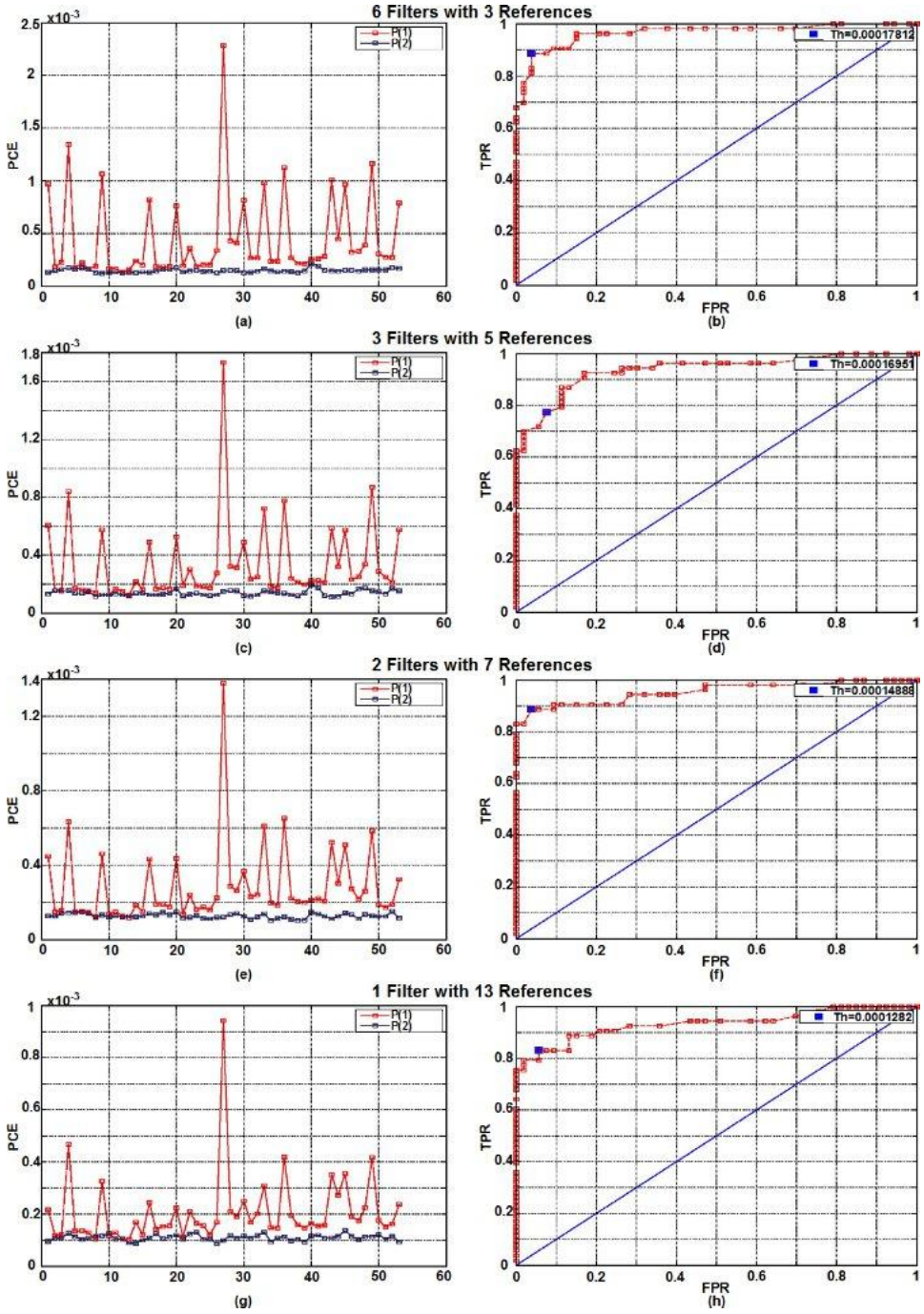


Figure. 22 PCE results and ROC curves using a segmented composite filter for 6 filters and 3-reference filters (a-b), 3 filters and 5-reference filters (c-d), 2 filters and 7-reference filters (e-f), 1 filter and 13-reference filters (g-h).

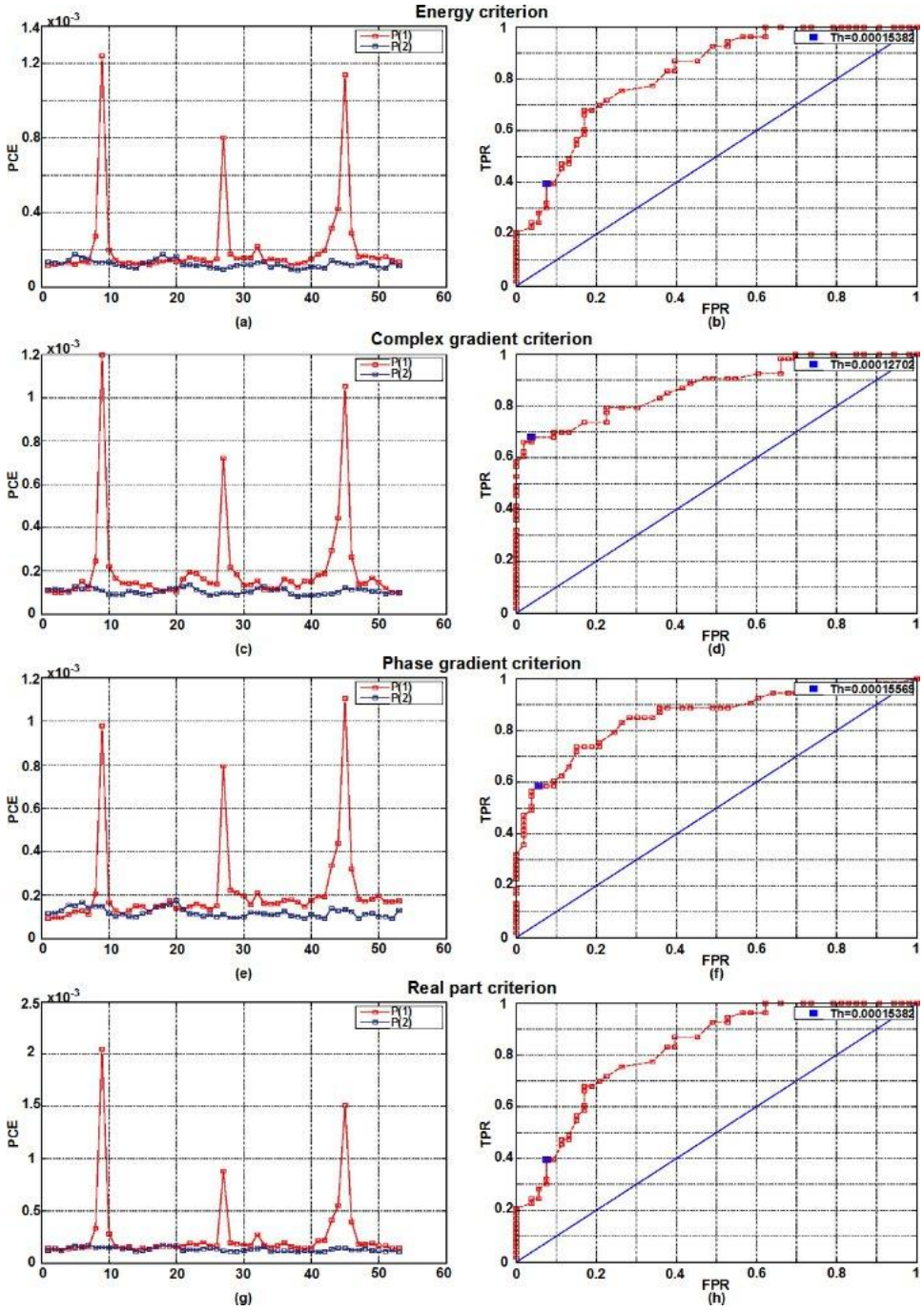


Figure. 23 PCE results and ROC curves using a segmented composite filter for the energy criterion filter (a-b), the complex gradient criterion (c-d), the phase gradient criterion (e-f), and the real part criterion (g-h).

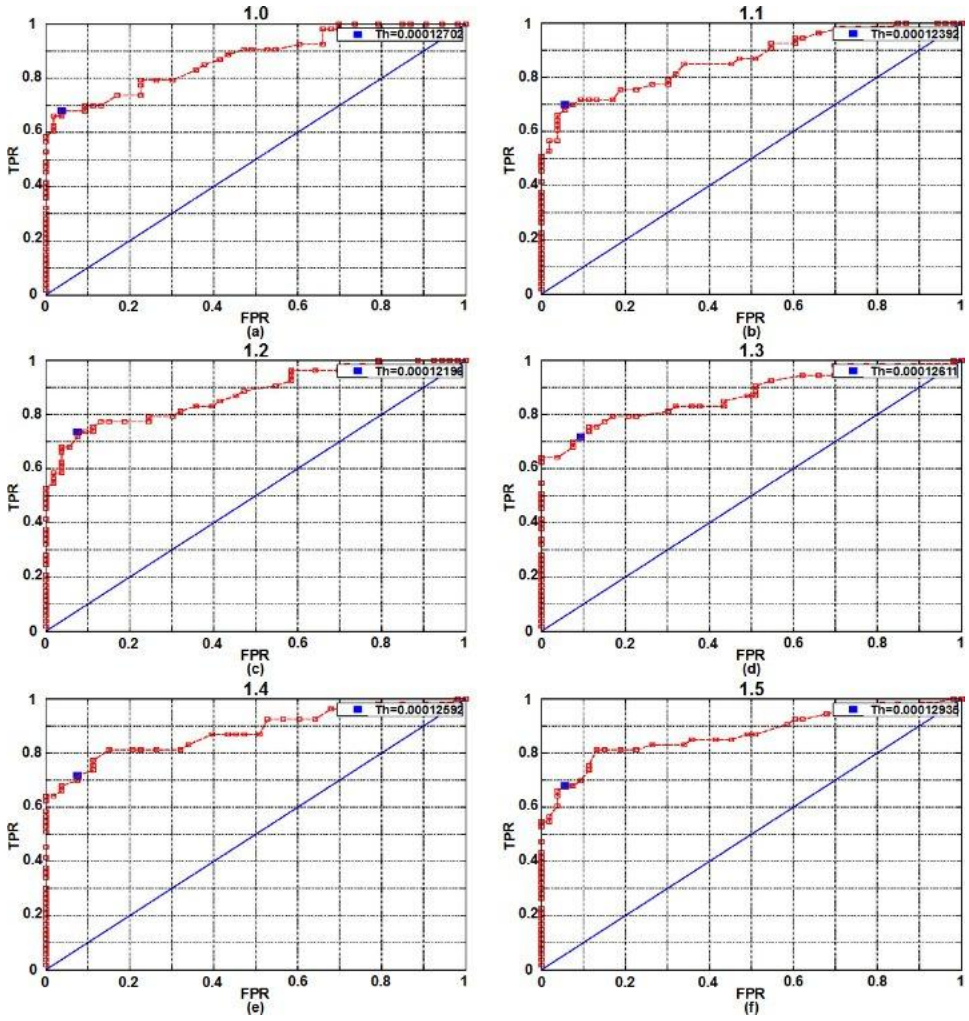


Figure. 24 Results using a segmented composite filter with different weight coefficients: (a) 1.0, (b) 1.1, (c) 1.2, (c) 1.3, (d) 1.4, (e) 1.5, and (f) 1.6.

As mentioned earlier, different segmentation criteria can be used for the fabrication of the segmented composite filter, i.e. energy of the spectrum (Figure 23 (ab)), its complex gradient (Figure 23 (cd)), its phase gradient (Figure23 (ef)), and its real part (Figure 23 (gh)). It is seen that the segmentation criterion has a significant impact on the performance of the filter, as exemplified by calculations of the recognition rates which are respectively equal (for $FPR = 0\%$) to 22% for the energy, 60% for the complex gradient, 32% for the phase gradient, and 21% for the real part. If the FPR is set to 10%, this recognition rate is respectively of 40%, 70%, 59%, and 39%. Although the phase gradient has a low TPR for FPR set to 10% its recognition rate is largely increased when the FPR is set to 0% contrasting with the energy and real part criteria. This is due to the fact that the phase contains most of the information. The most efficient criterion is that of the complex gradient leading to high

robustness and discrimination. The results obtained with the complex gradient are similar to those obtained with the energy criterion and 13 reference images.

As also indicated above, a weighting factor may be applied to some references allowing us to increase the face recognition performances. Several weighting factors were considered, when using a 3-reference filter (images 9, 27, and 45) segmented with the gradient spectrum. The effect is illustrated in Figure 26 with respective weighting factor 1.0 (Figure 24 (a)), 1.1 (Figure24 (b)), 1.2 (Figure 24 (c)), 1.3 (Figure24 (d)), 1.4 (Figure 24 (e)), and 1.5 (Figure 24 (f)). This weighting factor plays a significant role in the recognition performance. Comparing the TPR with a FPR set to 0% we pass from 51% (weighting factor=1.1) to 64% (weighting factor=1.3 and 1.4). When the FPR is set to 10% the TPR values are close to 70%.

COMPARISON

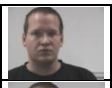
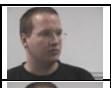
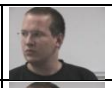
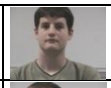
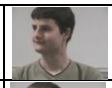
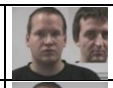
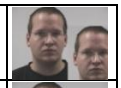
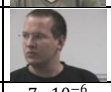
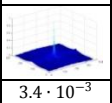
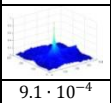
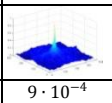
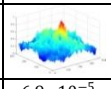
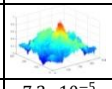
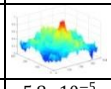
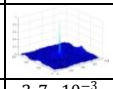
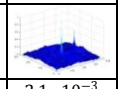
Finally, we turn to a comparison of the performances of the ICA (Sec. 3) and VLC (Sec. 4) methods focusing our attention on the role of the number and choice of the reference images, and the effect of the background noise of the target image.

Table 5 compares the results obtained via the two methods on the effects of the background noise of the image and the presence of a second face in the target image. The first line shows the target images used; the second line shows the reference images with minimal error obtained with the ICA-based method; the third line contains the error obtained; the fourth line shows the correlation plane, and the fifth line indicates the PCE value obtained with VLC-based method. For the ICA-based method, the three first images of the first line of Table 5 (subject 1) were used. The filter used for the VLC-method considers the first image of Table 5. As mentioned above, the ICA-based method allows us to recognize efficiently the subject when the background of the image does not contain any face. The 3 first images of Table 5, corresponding to the reference subject, lead to an error of the order of 10^{-31}, while the next 3 ones, corresponding to the second subject, lead to an error of the order of 10-6. The POF lead to very intense correlation peaks for the first 3 images, allowing a good recognition of the subject, i.e. the PCE values are of the order of 10^{-3} in case of recognition and 10^{-6} otherwise, while the errors of the ICA-based method are respectively of the order of 10^{-31} and 10^{-6}. While both methods have good recognition results, the ICA-based method leads to a more efficient discrimination. Now, as concerns the last 2 target images of Table 5 in which a second face has been introduced in its background, it is interesting to observe that the ICA-based method cannot recognize the subject in sharp contrast with the VLC-based method, i.e. the error is of the order of 10^{-6}, the POF has a high magnitude correlation peak, and the PCE value is of the order of 10^{-3} in a similar way when the reference subject is taken as the target image. Remarkably, VLC allows us to obtain results with much stronger robustness.

Finally, Figure 25 shows the results for both approaches. For VLC we use two segmented composite filters with 7 references, the complex gradient of the spectrum criterion, and a weighting factor of 1.4. The ICA-based method uses a single reference, i.e. image 27 (Figure 18). We note that these two methods present significantly different performances. For example, the ROC curve for VLC shows a good classification, with a TPR of respectively 90% and 97% when the FPR is set to 0% and 8%. In contrast, the ICA-based method leads to

TPR values of respectively 39% and 64% when the FPR is set to 0% and 8%. Hence, VLC leads to the best results with a good classification with only 2 correlations. The fact that this architecture is optically implementable makes this technique well suited for image processing systems since it is quasi-instantaneous, contrasting with the ICA method which is computationally intensive.

Table 5 : ICA errors and correlation planes using target images shown in the first row. RIME means reference image with minimal error.

Target image								
RIME								
Error	$1.3 \cdot 10^{-31}$	$1.3 \cdot 10^{-31}$	$2.7 \cdot 10^{-33}$	$7 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$7.3 \cdot 10^{-6}$	$2.7 \cdot 10^{-6}$	$1.1 \cdot 10^{-6}$
POF								
PCE	$3.4 \cdot 10^{-3}$	$9.1 \cdot 10^{-4}$	$9 \cdot 10^{-4}$	$6.9 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$	$3.7 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$

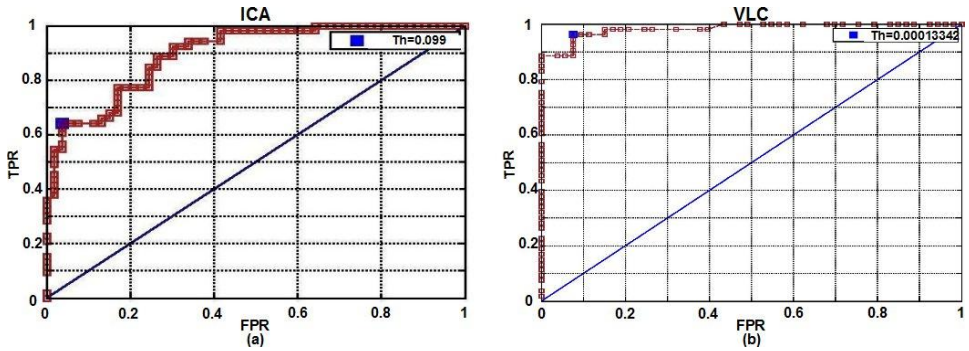


Figure. 25 ROC curves using 13 references for (a) the ICA-based method, and (b) the VLC method using a segmented composite filter.

SUMMARY

In this chapter, we proposed and validated a novel ICA-based approach for face recognition. The performance of the proposed techniques was compared with that of alternate techniques, such as the VLC. We have reported an extensive series of first principles numerical studies aimed at better understanding the role of the number and the choice of reference images. As mentioned earlier, one of the motivations of our study is to consider the effects of different criteria on the correlation peak detection and segmentation of the reference images for assessing the performance of the VLC. Additionally, we described a performance probabilistic metric for our classifiers i.e., the receiver operating characteristic (ROC). These two methods show different performances in terms of robustness and discrimination.

Importantly, the ICA-based technique possesses the best discriminating power, i.e. errors are on the order of 10^{-31} for the comparison of the subject with itself, and on the order of 10^{-6} for the comparison with a different subject. By comparison, the VLC-based technique is robust enough to treat the background noise of images. From the previous discussion we conclude that this technique is the most efficient for recognition of the entire database of 53 images for each subject. Hence, the VLC allows us to get good classification performances, i.e. typically 90% of true recognition for 0% false alarm with a segmented composite filter using the gradient of spectrum and a weighting factor of 1.4, while the ICA-based method offers weaker results. These findings may initiate future research efforts to improve the robustness of face recognition methods involving extremely large databases

REFERENCES

[1] P. Glascock and D. M. Kutzik. Behavioral telemedicine: A new approach to the continuous nonintrusive monitoring of activities of daily living. *Telemedicine Journal*, 6(1):33–44, 2004.

[2] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[3] L. Shen and L. Bai. A review on gabor wavelets for face recognition. Pattern Analysis and Applications, 9(2):273–292, 2006.

[4] T. Joliffe. Principal component analysis. New York: *Springer-Verlag*, 1986.

[5] P. Comon. Independent component analysis, a new concept? Signal Processing, 36(3):287–314, 1994.

[6] Alfalou and C. Brosseau. Robust and discriminating method for face recognition based on correlation technique and independent component analysis model. *Optics Letters*, 36(5):645–647, 2011.

[7] M. Barret. Traitement statistique du signal. 2009.

[8] Alfalou and A. Mansour. Double random phase encryption scheme to multiplex and simultaneous encode multiple images. *Applied Optics*, 48(31):5933–5947, 2009.

[9] VanderLugt. Signal detection by complex spatial filtering. *IEEE* Journals, 10:139–145, 1964.

[10] V. Oppenheim and J. S. Lim. The importance of phase in signals. Proceedings of the IEEE, 69(5):529–541, 1981.

[11] Alfalou, G. Keryer, and J.-L. de Bougrenet de la Tocnaye. Optical implementation of segmented composite filtering. *Applied Optics*, 38(29):6129–6135, 1999.

[12] V. K. V. Kumar and L. Hassebrook. Performance measures for correlation filters. Applied Optics, 29(20):2997–3006, 1990.

[13] J. Horner. Metrics for assessing pattern-recognition performance. Applied Optics, 31(2):165–166, 1992.

[14] F. M. Dickey and L. A. Romero. Dual optimality of the phase-only filter. Optics Letters, 14(1):4–5, 1989.

[15] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic roc curve. Radiology, 143(1):29–36, 1982.

[16] M. A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In University of Michigan, International Computer Music Conference, 2000.

[17] S. Makeig, M. Westerfield, T.-P. Jung, J. Covington, J. Townsend, T.J. Sejnowski, and E. Courchesne. Functionally independent components of the late positive event-related potential during visual spatial attention. The *Journal of Neuroscience*, 19(7):2665–2680, 1999.

[18] J. Onton, M. Westerfield, J. Townsend, and S. Makeig. Imaging human eeg dynamics using independent component analysis. Neuroscience and Biobehavioral Reviews, 30(6):808–822, 2006.

[19] M. Wang and Y. Mo. Face recognition method based on independent component analysis and by neural network. Proceedings of the SPIE, volume 5267, pages 214–219. SPIE, 2003.

[20] Alfalou and A. Mansour. All-optical video-image encryption with enforced security level using independent component analysis. *Journal of Optics A: Pure and Applied Optics*, 9(10):787, 2007.

[21] Alfalou, M. Farhat, and A. Mansour. Independent component analysis based approach to biometric recognition. In Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on DOI - 10.1109/ICTTA.2008.4530111, pages 1–4, 2008.

[22] X. Guan, H. H. Szu, and Z. Markowitz. Local ICA for the most wanted face recognition. In H. Szu Harold, Vetterli Martin, J. Campbell William, and R. Buss James, editors, Proceedings of the SPIE, volume 4056, pages 539–551. SPIE, 2000.

[23] Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[24] N. Gourier, D. Hall, and J.L. Crowley. Estimating face orientation from robust detection of salient facial structures. In Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures, 2004.

Joint Transform Correlation for face tracking: elderly fall detection application

Philippe Katz, Michael Aron, Ayman Alfalou

Vision_L@BISEN: Institut Supérieur de l'Electronique et du Numérique
20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France.

ABSTRACT

In this paper, an iterative tracking algorithm based on a non-linear JTC (Joint Transform Correlator) architecture and enhanced by a digital image processing method is proposed and validated. This algorithm is based on the computation of a correlation plane where the reference image is updated at each frame. For that purpose, we use the JTC technique in real time to track a patient (target image) in a room fitted with a video camera. The correlation plane is used to localize the target image in the current video frame (frame i). Then, the reference image to be exploited in the next frame (frame $i+1$) is updated according to the previous one (frame i). In an effort to validate our algorithm, our work is divided into two parts: (i) a large study based on different sequences with several situations and different JTC parameters is achieved in order to quantify their effects on the tracking performances (decimation, non-linearity coefficient, size of the correlation plane, size of the region of interest...). (ii) the tracking algorithm is integrated into an application of elderly fall detection. The first reference image is a face detected by means of Haar descriptors, and then localized into the new video image thanks to our tracking method. In order to avoid a bad update of the reference frame, a method based on a comparison of image intensity histograms is proposed and integrated in our algorithm. This step ensures a robust tracking of the reference frame. This article focuses on face tracking step optimisation and evalutation. A supplementary step of fall detection, based on vertical acceleration and position, will be added and studied in further work.

Keywords: correlation, joint transform correlator, elderly fall detection, video tracking, histogram similarity measure

1. INTRODUCTION

Because of the parallel evolution of life expectancy increase and falling birth rate, most of industrialized countries suffer from an ageing population. In order to address this problem, it is essential to find out new methods of elderly care. Indeed, traditional solutions (e.g. retirement homes) are no longer able to satisfy the increasing need. Consequently, it would be desirable to be likely to keep the healthiest part of this population at home. Firstly, in order to keep people's living standarts, habits and social relationships. Secondly, because a loss of familiar aspects and routines is a factor of old age diseases acceleration. However, keeping dependent people at home involves undoubtly new issues that have to be addressed. Indeed, old age diseases could lead to problems like therapy omission, frequent falls...

Within this new framework the "smart home" concept emerged and became a flourishing research field.[1–4] Demiris et al.[1] describe the term "smart home" as *"a residence equipped with technology that facilitates monitoring of residents and/or for promotes independance and increases residents' quality of life."* More precisely, it refers to a large set of elderly care applications, ranging from physiological and functional monitoring,[5,6] therapy delivering[7] and emergency situations detection and response.[8–17] Falls being one of the most serious brakes of keeping elderly at home, some solutions have been marketed, using either alert push-buttons or accelerometres.[18,19] In the first case the patient needs to be conscious and the system has to be worn in both cases. Thus, it may be impossible for the people in trouble to actuate this kind of device. Therefore, it becomes prominent to develop non intrusive systems. That is to say to design not wearable and no patient-actionated solutions. For that purpose, various types of sensors may be used, for example video cameras,[9–13,15] microphones[15,16] or smart floor.[13]

Because of the tedious installation process required for smart floors and the rich information of video cameras compared to microphones, video cameras seem to be a good compromise. In a simplistic way, a fall can be defined by a fast transition from a standing motionless to a lying down position. Hence, we propose to use face tracking methods to take into account temporal and video informations. In order to simultaneously compute the patient identification, detection and localization,

Optical Pattern Recognition XXIV, edited by David Casasent, Tien-Hsin Chao, Proc. of SPIE
Vol. 8748, 87480I · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2016413

we introduce in this article a tracking algorithm based on correlation and enhanced by a digital image processing method. Furthermore, we propose a validation protocol for our tracking system.

In this paper, we begin by presenting a short review of existing smart homes systems for elderly fall detection. We then justify and introduce a new iterative tracking method based on JTC (Joint Transform Correlator) correlation.[20] So as to validate our approach, an experimental protocol has been developed, allowing us to analyze the considered JTC parameters impacts on tracking performance. Experimental results, processed in a simulation room, are presented to validate our algorithm relevance.

2. RELATED WORKS

Various types of systems have been proposed to address the problem of elderly fall detection, using wearable (e.g .alert push-button,[17] accelerometres[10,11,17] or RFID tags[10,11,14]) and non-wearable sensors (e.g. video cameras[9–13,15] or microphones[15,16]). Särelä et al.[17] propose a system containing a wrist and a central station, called "IST Vivago®". By means of an accelerometre, they are able to detect falls and to analyze patient's muscular activity (making accessible his circadian rythm). As a fallback position, in case of non-detection of a fall, the wrist is fitted with an alert push-button. The main problem of their approach resides in the use of wearable sensors, that may be forgotten by the patient. Similarly, the "Smart Home Care Network"[10,11] also performs the detection process using an accelerometre. The position is recorded by an RFID tag worn by the patient. When a fall is detected, a video camera computes a posture analysis in order to reduce the false alarms rate. The good detection rate of patient's posture of this system is of 90%. Unfortunately, this approach is validated on a set of only 16 images which is unsufficient to cover the whole possible situations and environmental problems (e.g. edge, background or illumination). In the method proposed by Demongeot et al.,[8] called "Health Integrated Smart Home Information System", accelerometres, infrared sensors and magnetic door sensors informations are fused for fall detection. As with the previously described systems, this approach has the major drawback to be based on wearable sensors. In the perspective of having a non-intrusive, transparent solution, different methods using video cameras to monitor people's movements have been developed. Miaou et al.[21] use a system including a video camera and a convex mirror giving wide-angle pictures, the "SmartCam". Hung from the ceiling, their camera is able to record images from the whole room. The patient detection is performed by means of background substraction. Despite of the ability to obtain the whole scene using one single camera, some situations may be misinterpreted, like occlusions and quick sitting down. Williams et al.[9] propose a low-power distributed video cameras network for fall detection and patient localization. In order to reduce the image frequency, each camera node is sampling at a rate lower than $1/5Hz$. The person is detected each frame and the fall is detected using a simple threshold applied on the aspect ratio of the patient. Once a fall is detected, the person is localized by means of homography estimation between each synchronized node of the distributed network. Their algorithm yields 95% of fall detection good classification rate, using a set of only 40 images. Such as the "Smart Home Care Network", the results seem to be performed using an unsuficient set of images. Furthermore and as specified by the authors, this approach may be inaccurate for complex fall situations (e.g. falling on a couch), on which temporal information may be essential. De Silva et al.[15] combine information from microphones and video cameras to detect emergency situations (falls, therapy omission, shouts). For that purpose, the patient detection is performed by means of background substraction. The background is updated each frame according to the person location. A shape/shoulder model is then applyied on the subject, the corresponding trunk being indentified through a projection matching procedure. The tracking process is realized with a histogram matching, a state-based approach making possible the events recognition. Finally, the audio event recognition is done by a pitch contour detection. The video and audio based fall detection accuracy are 93.3% and 91.67%, respectively. Regrettably the authors do not give any time performance of their method. No information is also available on how they fuse the data extracted from audio and video sensors.

The main problem of some of those methods[8,10,11,17] resides in the use of wearable sensors, as it can be forgotten by a patient suffering from a memory-loss desease. Consequently, our method is mainly based on video sensors, that yields rich information and make assessible the largest variety of falling situations. While Williams et al.[9] do not take into account the temporal information, which may be necessary for complex events analyzis, we propose to perform a patient's face real-time tracking by means of a low-resolution video camera. There are many available person detection algorithms in literature,[22] like point detection (e.g. Scale Invariant Feature Transform[23]), segmentation (e.g. graph-cut,[24] active contours[25]), background modeling (e.g. Mixture of Gaussians[26]) or supervised classifier (e.g. Support Vector Machines[27]). For the sake of taking advantages of the simultaneous identification, detection and tracking process abilities of correlation and of figuring out its reliability in such application, we propose a method based on a Joint Transform Correlator[20] (JTC).

Lastly, our algorithm's performances, namely its detection precision and the computational time, are evaluated on a large set of video sequences.

3. CORRELATION METHOD: TRACKING

The principle of correlation, consisting in a simplistic way to compare a target image (image to be recognized) to a reference image (from a database), has aroused much interest among researchers all over the world.[28] Two principal correlation architectures may be found in the litterature: the JTC (Joint Transform Correlator[20]) and the VLC (VanderLugt Correlator[29]). Both architectures are based on the classical "4f" set-up, composed of an entry plane, a first Fourier transform (FT), a Fourier plane, a second FT and an output plane, or correlation plane (FT^{-1} of the Fourier plane). The principal difference of those architectures is a consequence of the different process applyied on input and Fourier planes.[28,30,31] In this paper, we only focus on JTC correlator. Our interest on correlation is motivated by: (i) the relative algorithm simpleness (i.e. spectral comparison between the target and reference images); (ii) the global processing of the target image; (iii) the simultaneity between the detection and localization stages.

The scientific output on correlation field ran out of slackened off among the past few years. It can be explained by three principal reasons. Firstly, tremendous effort were done on proposition and validation of correlation filters,[28,30,31] i.e. the Fourier plane, at the expense of entry and correlation processing. Indeed, we have recently showed that some specific process on the correlation plane can yield to substantial improvements of correlator performance.[31] Furthermore, we have proposed and validated a correlation plane denoising method,[31] leading us to enhance the correlator decision-making abilities: increase of good recognition rate and decrease of false positive. Secondly, the dissemination of correlation is also dued to an emphasis on an all-optical implementation.[28] Without any doubt, this, with all the problems generated by optical set-ups, complicates the use of this method. Moreover, in many situations, the video frequency if fastly enough (using programmable units, like GPUs[32]). To conclude, our opinion is that correlation has to be considered as a baseline method. Consequently it has to be completed with others methods of data fusion and image processing. Here, we propose to use and optimise a JTC-based face tracking algorithm, allowing the addition of a fall detection step subsequently.

3.1 Joint Transform Correlator: principles

Initially introduced by Weaver and Goodman,[20] the Joint Transform Correlation is an optical recognition method based on the combined presence of the target (image to be recognized) and reference images on a entry plane. The algorithm yields a correlation plane, presenting two correlation peaks which intensities being dependant of images similarity. The correlator is schematically presented on figure 1.

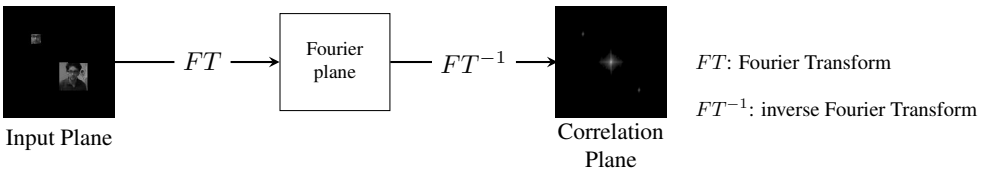


Figure 1: Joint Transform Correlator.

The method consists on the following items: (i) a Fourier transform (FT) of the input plane yields the Fourier plane $|t_{(u,v)}|$ (Eq. 1); (ii) the square modulus of this first tranform is performed $E_{(u,v)} = |t_{(u,v)}|^2$; (iii) finally, an inverse Fourier transform (FT^{-1}) returns the correlation plane. The correlation plane contains three main peaks: a central peak, or "zero-th order", corresponding to the sum of each images' autocorrelations (FT^{-1} of the first two term of equation 1) and two peripheral peaks, corresponding to the correlation between the the target and reference images (FT^{-1} of the two last terms of equation 1).

$$E_{(u,v)} = |t_{(u,v)}|^2 = \begin{cases} |S_{(u,v)}|^2 + |R_{(u,v)}|^2 \\ +|S_{(u,v)}|e^{\phi_s(u,v)}|R_{(u,v)}|e^{-\phi_r(u,v)+j(ul+vl)} \\ +|S_{(u,v)}|e^{-\phi_s(u,v)}|R_{(u,v)}|e^{\phi_r(u,v)-j(ul+vl)} \end{cases} \quad (1)$$

- $E(t)_{(u,v)}$ *is the resulting correlation plane,*
- $t_{(u,v)}$ *is the joint spectrum,*
- $|S_{(u,v)}|$ *and* $|R_{(u,v)|}$ *are target and reference amplitude spectrum, respectively,*
- $\phi_s(u,v)$ *and* $\phi_r(u,v)$ *are the target and reference phase, respectively,*
- u *and* v *are the pixel coordinates,*
- l *is the distance between target and reference images in the input plane.*

The correlation peaks intensity is conditioned by the degree of similarity between the target and reference images. Concerning to their locations, they are dependent on the images relative location on the input plane. Those two properties are essential for the object detection and localization. Knowing the reference image's location and reference and target image's relative location, it is thus possible to perform an object tracking. Unfortunately, this classical correlator suffers from two major drawbacks: the zero-th order, which is much more intense with respect to the correlation peaks, and the two low-intensity and large correlation peaks, making approximative their localization. As will be shown in part 4.1.4, the input plane's size impacts in opposite ways both the computational time and the object detection accuracy. Thus, reducing the input plane is only possible if another manner of improving the JTC correlation is found. The classical JTC being limited by the two problems previously exposed, it is necessary to cancel out the zero-th order. For that, we need to remove the first two terms of equation 1 (i.e. $|S_{(u,v)}|^2 + |R_{(u,v)}|^2$). This can be done first by creating two supplementary input planes presenting separately the target and reference images and then by deducting independently their computed Fourier transform from the joint spectrum (Fig. 2).

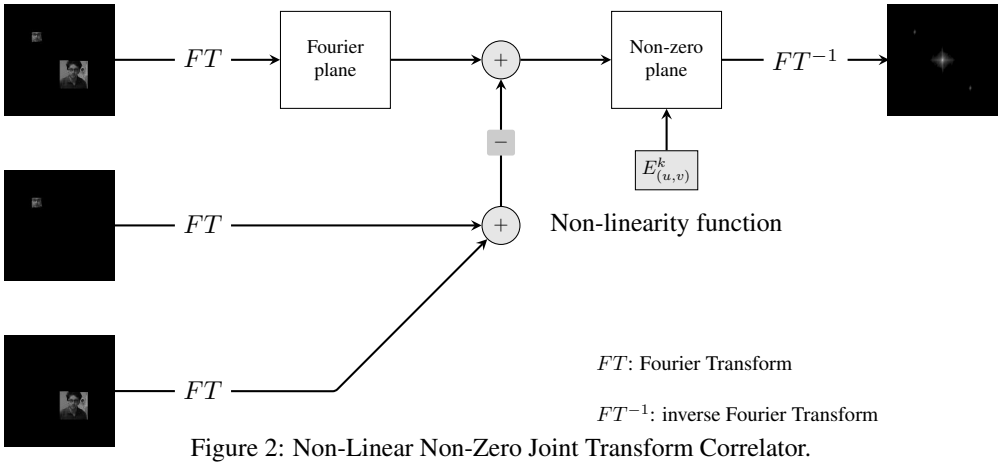


Figure 2: Non-Linear Non-Zero Joint Transform Correlator.

Lastly, in order to address the problem of large correlation peaks' low-intensity, Javidi[33] proposes to introduce a non-linearity function in the Fourier plane (i.e. by means of a coefficient of non-linearity $k < 1$, $E^k_{(u,v)}$, Fig. 2). The value of k highly influences the JTC perfomance. Indeed, using a k near to 1 will obviously lead to low-intensity and large correlation peaks whereas those peaks will be sharp and intense with k near to 0. This will consequently affect the JTC behaviour: large correlation peaks increases the JTC robustness at the expense of its discrimination and its location precision. A compromise between robustness on the one side and discrimination and location precision on the other side has thus to be found. A study of the non-linearity coefficient effect is performed on part 4.1.1.

3.2 Correlation applied to face tracking

The peaks' location being dependent on the relative location of the images in the entry plane[34] (Eq. 1), knowing the location of the reference image, it becomes possible to obtain the tracked object's location in our target image. Thus, our algorithm is based on an iterative method: the detected region in the current target image (time t) becomes the reference image at time $t+1$, the correlation allowing to perform the detection and tracking in a same process. Concretely, the nz-nl-JTC (Non-Zero Non-Linear JTC) is initialized by a reference image, placed in the input plane's top-left panel. A target image is then inserted in the opposite panel. The JTC process yields a correlation plane, which most powerfull peaks, namelly the correlation peaks, are used for object (i.e. the object displayed by the reference image) localization. This detected region is then introduced as a reference image for the next iteration, the target image being the following video frame. This

self-adapting architecure is consequently able to track a moving object with shape transformation or a changing point of view.

The perspective of our tracking algorithm is to perform an elderly fall detection. Hence, we propose to use the patient's face as it is the richest and the less clothes dependent part of a person. For that purpose, the initialization stage is then completed by a Viola and Jones' cascade classifier.[35] The training set is performed using 5000 frontal images (positives) and 3000 negatives examples.[36]

Decimation As previously said, it is imperative to decrease the computational time. Therefore we propose to study the effect of another tracking parameter on performances, namely the decimation. More precisely it consists to take into account a lower number of frames for the tracking by reducing the frame frequency. A decimation $d = n$, n being an interger, will lead to take into consideration only $1/n$ frames from the total amount of images. This parameter is studied on part 4.1.3.

3.2.1 Video sequences

Seeking for an extensive experimentation of our algorithm, we define an experimental protocol exploring the tracking limitations. Thus, we set aside a special room (Fig. 3), reproducting an hospital or retirement house room. Furthermore, a large variety of scenarii have been imagined. Firstly, a simple sequence of 535 images have been recorded (resolution: 640×480, 30 frames per second). This sequence, used to access the effects of JTC parameters and the further need of optimisations, comprises two falls. The first fall's onset occures frame 13 and lasts approximately $1s$. Regarding the second fall, lasting around $200ms$, it begins frame 106.

Figure 3: Simulation room.

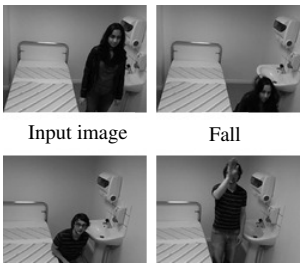


Figure 4: Some experiment situations (falls and occlusion) used to test our algorithm.

Secondly, we created a broader set of scenarii in order to observe our algorithm accuracy. This set contains 14 different events making possible to isolate the binding cases of tracking and sort them by situation. The different events are shown on table 1. They are organized by types of movements (none, i.e. standing position, face rotation on the whole directions, translation, occlusions, or when the tracked person is out of scope), by speed (fast or slow), by amplitude (soft or high) and direction (back, front, up or down). The figure 4 illustrates some of these events, namely a fall (images on top and in the bottom left) and an occlusion (bottom right), corresponding to events $\#5$ or $\#7$ and $\#13$, respectively. Five different people and from different ethnical origins where registered. The database is composed by 21087 frames. The face region position on each frame have been manually recorded, leading a "field-reality tracking". The distance between the corresponding regions' positions extracted by means of manual and JTC tracking have been computed to observe our algorithm precision and eventual loss of tracking. The face region is fixed for tracking methods at a square of side $78px$.

3.2.2 Tracking optimisation

Unfortunately, the JTC method (described Fig. 2) suffers from two major drawbacks. First of all, in case of low similarity between target and reference images (e.g. blurred image because of a fast falling patient), the JTC yields low-intensities correlation peaks. Hence the system becomes noise-sensitive. This perturbation is likely to occur as we are using a low-resolution video camera to spare computational time. Second, the localization in the current image (target image) only depending on the previous frame localization (reference image), a loss of tracking (e.g. a sudden focus on background

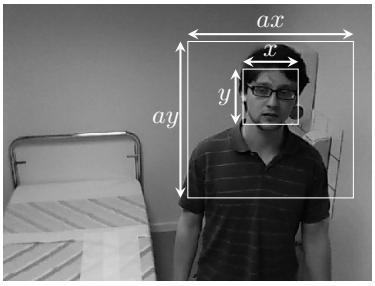
Table 1: Experimented situations.

Event					Number of images
#	Movement	Speed	Amplitude	Direction	
1	None				12 496
2	Face rotation	Slow	Soft		1 391
3	Translation	Slow		Back	1 380
4	Translation	Slow		Front	895
5	Translation	Slow		Down	358
6	Translation	Slow		Up	318
7	Translation	Fast		Down	152
8	Translation	Fast		Up	341
9	Face rotation	Fast	Soft		860
10	Face rotation	Fast	High		986
11	Translation	Fast		Back	578
12	Translation	Fast		Front	318
13	Occlusion				679
14	Out of scope				335
Total					21 087

image) cannot be detected or corrected by an iterative algorithm. Consequently, this kind of techniques makes definitive any loss of tracking.

To remedy that, we propose two optimisations. The first one is based on determinist information: the face on current target image may be present on a specific and localized region. According to that, the correlation computing is only usefull on this precise region. The second correlation is done by calculating a histogram similarity measure between the reference image and the detected face region on target image. This supplementary step allows a loss of tracking detection and hence an algorithm reinitialization.

Region of interest definition We introduce a first optimisation by reducing the research of patient's face to a limited region of interest. To achieve this, we propose to use prior-knowledge on possible face location. More precisely, a patient's location may only be a narrow field, all the more while looking for disabled or elderly people. Consequently, we define a region of interest in the target image arround the face region detected in the previous frame. As shown on figure 5, a scale factor a is applied on the previous face region $s = (x, y)px$ to obtain a region of interest $s_{roi} = (a \times s)px$ that will be finally used as a target image for the correlation. As we are working on an eldely fall detection application, we are mostly interested by the lower part of the image, under the patient's face. Thus, the region of interest is biaised to the image bottom. We still keep a small part over the patient for some unexpected situations. A study of the size of this region is presented on part 4.1.2.



x and y: the size of the detected face region

a: the scaling factor

Figure 5: Research region definition.

Histogram similarity correction As previously said, the major drawback of iteratives algorithms resides in their inability to detect when the tracking process is lost (e.g. when it keeps tracking the image background). To remedy that, we propose to use the information given by the histograms from reference image and detected region. Specifically, for an image an histogram is a tool of pixels' intensities repartition. Consequently, two histograms coming from a visually similar region will be fairly close whereas they will radically differ while coming from different image regions. Hence we quantify the similarity between the detected regions in successive frames by means of histogram comparison, that is the Pearson Chi Square.[37] As explained in equation 2, the Pearson Chi Square $X^2(H_{t-1}, H_t)$ is calculated by adding, for each intensity I, the squared difference between its probability density in the current (time t) and previous (time $t-1$) detected regions ($H_t(I)$ and $H_{t-1}(I)$, respectively) and divided by the one at previous frame $H_{t-1}(I)$. Finally, having the ability to detect a loss of tracking, it becomes possible to reinitialize our algorithm and hence having permanent knowledge of the patient's location. This reinitilization step is even more essential as an elderly fall application requires a continuous tracking without any external intervention.

$$X^2(H_{t-1}, H_t) = \sum_I \frac{(H_{t-1}(I) - H_t(I))^2}{H_{t-1}(I)} \tag{2}$$

- H_{t-1} *and* H_t *are the histograms computed from detected region at time* t *and* $t-1$*, respectively,*
- I *is the corresponding histogram bin (i.e. intensity).*

The figure 6 illustrates a loss of tracking situation. These results have been obtained using the 535 frames video sequence as described part 3.2.1 for decimation (see page 5) $d = 1$ (*solid lines*) and $d = 2$ (*dashed lines*), that is $30fps$ and $15fps$ respectively. The figure 6a presents the distance in pixels between the JTC tracking without histogram correction and the manual tracking. For the first fall (onset at image 13, $\sim 1s$), both curves are able to track accurately the patient, as the relative distance is lower than $78px$ (the face region side). Concerning the second fall (onset at image 106, $\sim 200ms$), we can clearly see a sharp change for $d = 2$, going from a distance of around $0px$ to more than $200px$, whereas the solid line, corresponding to $d = 1$, stays under than a distance of $50px$. This abrupt fluctuation corresponds to a loss of tracking situation. Another noticeable result is that the distance for a decimation of 2 is still fairly high compared to the case using $d = 1$. This is dued to the fact that once the tracking is lost, it is unable to reinitialize itself.

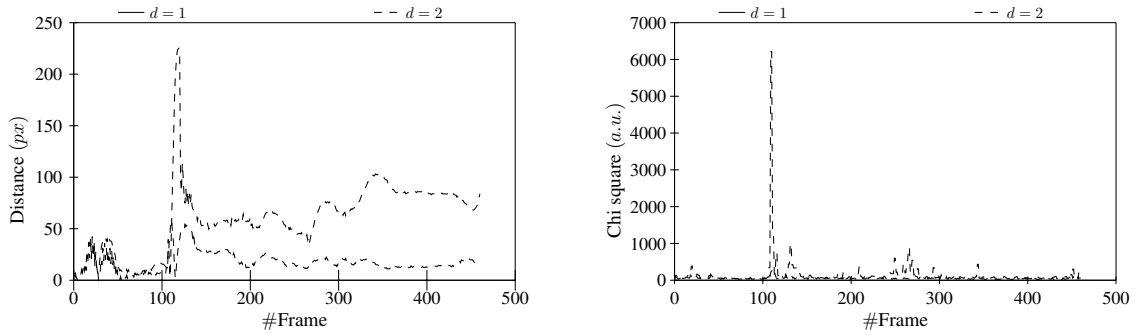
Consequently, it becomes essential to find a means of detecting such sitations. The figure 6b shows the histogram similarity measure, namely the Pearson Chi Square, as previously explained. The perceptible result that can be observed is the relatively low X^2 value in normal situation (approximately 0), that is to say, the lines corresponding to $d = 1$ for the whole sequence and to $d = 2$ before image 106 (on which the algorithm is loosing the patient). For $d = 2$ we observe a very sharp peak (> 6000) at image 106. Thus, the Pearson Chi Square similarity measure is a powerfull tool for loss of tracking detection. The line returns then to low values as it is now tracking a part of the background, the histograms of detected regions are hence kept unchanged. A study of the effects of histogram correction on tracking accuracy is given part 4.2.

4. EXPERIMENTAL RESULTS

In this part, we firstly describe the effect of tracking parameters on performance (part 4.1), namely: the non-linearity coefficient, the size of region of interest, the decimation and the size of correlation plane. The histogram optimization is evaluated in 4.2. Results obtained for each event described part 3.2.1 will be presented 4.2.

4.1 Effects of tracking parameters on performance

In order to optimize the JTC tracking performance, we study the effects of the different available parameters on accuracy and computational time. As desribed part 3.2.1, results have been performed by means of a 535 frames video sequence, containing two falls, one slow ($\sim 1s$) and one fast ($\sim 200ms$), begining frame 13 and 106, respectively. The figure 7 presents the relative distance in pixels between JTC tracking without histogram optimization (as we are studying only the tracking parameters) and the reality-field tracking. Concerning the figure 8, it illustrates the effects of decimation and correlation plane size on computational time per frame, in order to obtain a real-time tracking. Results have been obtained by an average on the whole sequence and on the whole resting parameters values.



(a) (b)

Figure 6: Loss of tracking: (a) distance between manual and JTC tracking and (b) Chi Square calculation in function of frame number for a decimation value of 1 (*solid line*) and 2 (*dashed line*) and using as parameters values $s_{plane} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $k = 0.4$.

4.1.1 Non-linearity coefficient k

Results presented on figure 7a correspond to the effects of non-linearity coefficient. For that purpose we performed the tracking process for k ranging from 0.1 to 0.9. For lisibility reasons, only representative results are illustrated here, that is $k \in [0.3; 0.6]$. For $k = 0.3$ and $k = 0.6$ we can observe that both lines are varying abruptly. This variation occurs at frame 13 and 106 for $k = 0.3$ and $k = 0.6$, respectively. The relative distance going from around $0px$ to $100px$ approximately. As the face region side is $78px$ long, the patient's face is lost by the tracking process. Indeed, none of these curves show a significative decrease after this event. This loss of tracking is explained by the noise sensibility of JTC using those values, correlation peaks being for those cases lower than subsidiary background induced peaks.

Concerning the coefficient values $k = 0.4$ and $k = 0.5$, they are substantively similar, curves on figure 7a being almost identical. We observe no brutal variation for those values, both allowing for this video sequence a tracking for the whole experiment duration. A comparison with other video sequences and an observation of resulting tracking images let us to experimentaly determine an optimal value of non-linearity coefficient $k = 0.4$.

4.1.2 Size of region of interest $\mathbf{s_{roi}}$

In a similar way, we observe on figure 7b the effects of the region of interest used as a target image. The region of research size is defined in term of scale factor applied on previously detected face region (here, $s = (78, 78)px$). While using a fairly small region of interest ($s_{roi} = (2 \times s)px$), the patient's movement between two frames may lead its face to be out of this region (large or fast movements). Regarding the use of a large region of interest ($s_{roi} = (4 \times s)px$), image's brackground induced peaks may be not efficiently limited. Indeed, the corresponding lines both present an abrupt distance variation at frame 106. Finally, a region of interest $s_{roi} = (3 \times s)px$ represent the best compromize between those two extreme cases, as it does not lead in this sequence to lose the tracking.

4.1.3 Decimation d

The figure 7d illustrates the results for different decimation values. Similarly to the non-linearity coefficient, the results have been realized for d ranging from 1 to 6 but the representative ones only are presented here. We can observe that an increase of decimation leads to a decrease of tracking accuracy. Indeed, while using the whole frames ($d = 1$), the tracking is not lost for this video sequence, whereas $d = 2$ and $d \geq 3$ generate a loss of tracking at frames 106 (fast fall) and 13 (slow fall), respectively. That can be explained by an insufficient frame rate for an accurate tracking. That is the difference between two successive frames is too large to obtain powerfull correlation peaks, leading to a loss of tracking.

The effects of decimation on computational time per frame are presented on figure 8a. We observe that the computanional time is predictabily increasing while reducing the decimation value, going from $0.8s$ for $d = 1$ to $0.14s$ for $d = 6$. As previously explained (Fig. 7c), the decimation highly affects the tracking accuracy. As a decimation upper than 3 does not reduce significatively the computational time per frame, we only focus on $d \in [1; 3]$ for part 4.2.

4.1.4 Size of correlation plane $\mathbf{s_{plane}}$

As we can observe on figure 7d, presenting the effects of correlation plane size going from $(128, 128)px$ to $(1024, 1024)px$, the best tracking results are obtained for a size of $(512, 512)px$. A larger size increases the noise in correlation plane, whereas a lower size reduces the localization precision. Hence it explains the loss of tracking effects observed for correlation plane size of $(128, 128)px$, $(256, 256)px$ and $(1024, 1024)px$.

Finally, the figure 8b illustrates the computational time per frame as a function of correlation plane size. We observe that the curve follows an exponential tendency. This is dued by the three Fourier transform processing, needed for the non-linear non-zero JTC architecture. Reducing the correlation plane is thus essential to process the tracking algorithm in real-time. A seen on figure 7d, our work progress does allow for the moment an accurate tracking with a $(256, 256)px$ correlation plane.

4.1.5 Conclusion

As explained, all of those parameters highly affect the tracking process, in terms of localization precision or of computational time. The non-linearity coefficient generates changes on correlation peaks shape, as a low value will lead to sharp and powerfull peaks and a high value to broad peaks. Hence a compromize have been found for $k = 0.4$. Likewise, the region of interest side will have to be large enough to comprise the patient's face in the following frame but also small enough to reduce significatively the noice induced by the image background. Concerning the decimation and the size of correlation plane, their effects are twofold, as they act both on localization precision and on computational time. Hereafter in this article, we use $k = 0.4$, $s_{plane} = (512 \times 512)px$ and $s_{roi} = (3 \times s)px$ as they are the parameters giving the bests results on JTC tracking in our testing room. This experimental protocol can be easily reproduced in another experimental configuration.

4.2 Effects of histogram correction

For the sake of an evaluation of our algorithm, we experiment in this part a comparison of the tracking accuracy of the JTC tracking with and without histogram correction. For that purpose we used the set of video sequences described part 3.2.1, generated from 5 subjects of different ethnical origins. The whole dataset, comprising 21087 images, contains 14 different events (see Tab. 1). The table 2 illustrates the results generated. Are presented on this table: the average distance between the JTC and the manual tracking (performed only on tracked frames, that is frames where the tracking is not lost) and the percentage of tracked and non tracked pictures for decimation values from 1 to 3. The histogram similarity measure have been processed for both JTC tracking cases, with and without histogram correction. For the algorithm without optimisation, frames have been automatically labeled as non-tracked while the Pearson Chi Square is $X^2 > 100$ and while the measured distance with the manual tracking is higher than $78px$ (as the face region is $78px$ side). For the JTC tracking with histogram optimisation, frames on which the patient is tracked by means whether the JTC or the Viola and Jones detector have been labeled as tracked pictures.

Firstly, we can observe that the average distance is in any case lower than 30. The face region occupying an area of $(78, 78)px$, the overall face localization went thus well. This distance is sightly higher for the non-histogram correction case. Indeed, an iterative tracking may lead to a slow deviation over time of the detected region according to the real face location. This deviation is usually corrected by the histogram optimisation. The average distance as a function of the decimation value is here given in an indicative manner as it is not a significant comparison criterion. Furthermore, we can see a meaningful increase of tracked pictures percentage while using the histogram correction. This improvement is of $21.61pts$, $23.35pts$ and $20.25pts$ for a decimation of 1, 2 and 3, respectively. Thus, the use of this supplementary step of histogram optimisation allows to highly improve the tracking process. Finally, we can observe that the higher detection rate occurs for $d = 2$. Indeed and as previously said, the tracking process may lead to a slow deviation of the detected region. The same deviation induces a more important change between two images while using $d = 2$ than $d = 1$. Hence, as the difference between two consecutive histograms is broader, a larger part of theses deviations will be addressed by the histogram opimisation with $d = 2$. In contrast, using of $d = 3$ generates a too low frame rate.

The results obtained for the 14 tracking situations (see part 3.2.1, Tab. 1) are presented in table 3. Some noticeable results may be observed, especially for borderline cases for iterative tracking, that is fast fall, fast and high face rotations, occlusions and while the patient is out of scope (events $\#7$, $\#10$, $\#13$ and $\#14$). First of all, we can see a significant improvement of detection rate (around $20pts$) in case of fast fall, and this for decimations of 1, 2 and 3. Hence, we get for that precise case a detection rate of 69.74% and 61.84% for $d = 1$ and $d = 2$, respectively. Furthermore, another noticeable

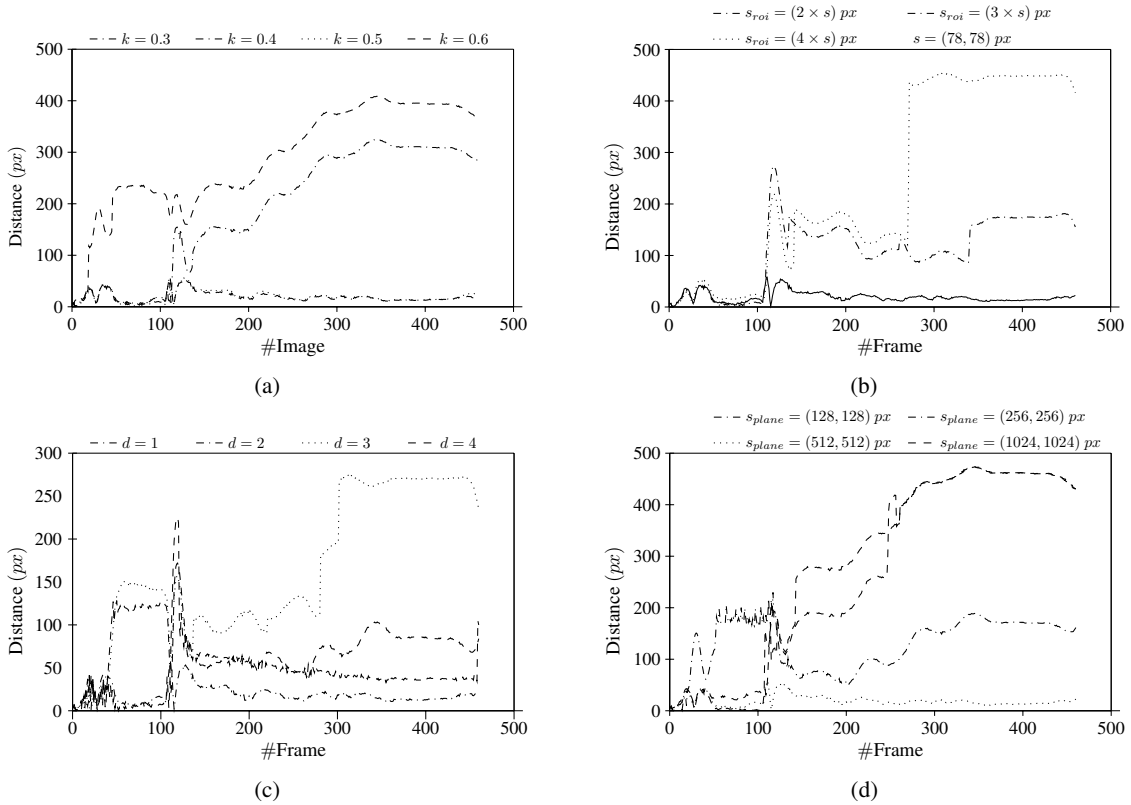


Figure 7: Distance between manual and JTC tracking in function of frame number: (a) for a coefficient of non-linearity $k \in [0.3; 0.6]$ and using as parameters values $s_{plane} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $d = 1$; (b) for a region of research $s_{roi} = (a \times s)px$, $a \in [2; 4]$, $s = (78, 78)px$ and with $s_{plane} = (512, 512)px$, $d = 1$ and $k = 0.4$; (c) for a decimation $d \in [1 : 4]$ and with $s_{plane} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $k = 0.4$; (d) for a correlation plane size $s_{plane} = (2^x, 2^x)px$, $x \in [7; 10]$ and with $s_{roi} = (3 \times s)px$, $s = (78, 78)px$, $d = 1$ and $k = 0.4$.

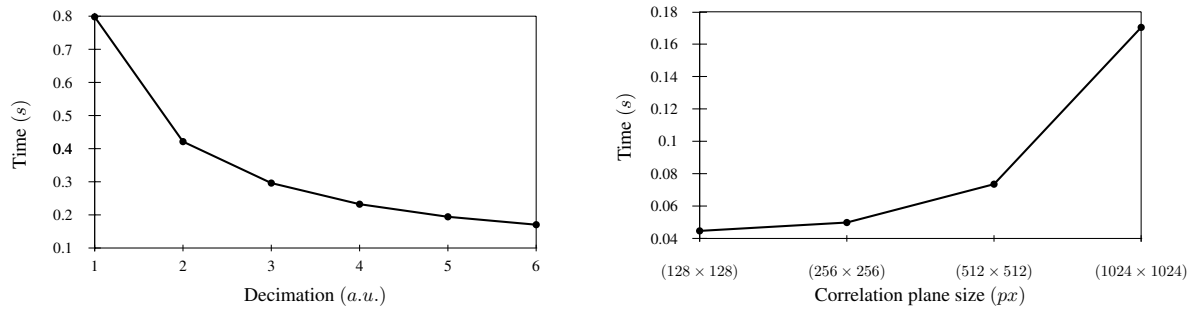


Figure 8: Average processing time ($k \in [0.1; 0.9]$, $s_{roi} = (3 \times s)px$ and $s = (78, 78)px$): (a) in function of decimation, using $s_{plane} = (512, 512)px$; (b) in function of correlation plane size, using $d = 1$. C++ / OpenCV, CPU Intel Core ir-2400 3.10 GHz, 4Go RAM, Windows 7 Enterprise 64 bits.

improvement occurs for fast and high amplitude face rotation (more than $51pts$ for $d = 2$). Concerning the occlusion and out of scope cases, the detection rate is increasing of $24.74pts$ and $24.47pts$ for $d = 1$ and $d = 2$, respectively. Lastly, we surprisingly observe some cases of bad detection while the patient is in static standing situation (event $\#1$). These cases are caused by loss of tracking occuring in previous events.

Thus, our optimised JTC tracking algorithm induces significant improvements, as it reduces the amount of non-detection rate. Indeed, as a loss of tracking was definitive, this histogram correction make the method able to detect such cases and thus to reinitialise itself. As the decimation increases the non-detection cases it becomes able to balance between JTC and Viola and Jones method and hence a compromised has to be found. This compromise is obtained for a decimation $d = 2$, taking the best advantages of both methods.

Table 2: Global comparison results in term of decimation and of presence of JTC correction ($s_{plane} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $k = 0.4$).

	Decimation					
	1		2		3	
Tracking / Value	Classical	Histogram	Classical	Histogram	Classical	Histogram
Average distance	26.21	16.11	20.34	18.46	19.86	20.26
Tracked pictures (%)	57.96	79.57	58.11	**81.46**	55.79	76.04
Non-tracked pictures (%)	42.14	20.43	41.89	18.54	44.21	23.96

5. CONCLUSION

This article presents (a first step of our elderly fall detection method) a detailled study of JTC tracking algorithm and evaluates the effects of parameters affecting its process. Furthermore, we propose two improvements of the iterative JTC tracking approach, that is a region of interest definition and an histogram comparison based optimisation. The use of this method ensures a real time face tracking by means of a simple low resolution webcam, that can be applied for an elderly fall detection in further work. An experimental protocol, using a significant amount of testing images, have been introduced, allowing to identify the best compromise of JTC parameters for a given configuration and to understand and analyze the behaviour of a our tracking algorithm for different possible borderline cases. In addition, as our experimental protocol is reproductible, it can be used to determine the best parameters configuration for various kinds of applications.

Supplementary improvements may be obtained while merging our algorithm with a skeleton detection method, in order to propose a detection not only of the patient's face but also of its entirely. It would be also interesting to take into account the depth information to adapt and optimise the region of research in our algorithm. Furthermore, the use of wide-angle pictures (MapCam[21]), posture detection methods[10,15] and audio data[15] are encouraging initiatives that can be fused with our algorithm. In doing this, our approach may be greatly enhanced, each method compensanting each other deficiencies.

Seeking for a validation of our algorithm, an evaluation study have to be performed by means of a larger number of individuals and situations. Moreover our approach have to be compared with standard video detection method widely explored in literature. As previously said, these methods may be based on point detection,[23] segmentation,[24,25,38] background modeling[26] or on supervised classifier.[27,39] This article proposes a improvement of JTC tracking method, merging different image processing algorithm. As the main aim of our approach is an elderly fall detection application, that can be embedded in a smart home environnement, the algorithm optimisation have been consequently adjusted and analysed. To achieve this study, a fall detection criterion has to be added on our method and validated. Lastly, a fall detection ratio has to be quantified on a large dataset and compared with results of other fall detection techniques.

REFERENCES

[1] Demiris, G. and Hensel, B. K., "Technologies for an aging society: a systematic review of "smart home" applications.," *Yearbook of Medical Information* **31**(0943-4747 (Linking)), 33–40 (2008).

[2] Chan, M., Estve, D., Escriba, C., and Campo, E., "A review of smart homes: Present state and future challenges," *Computer Methods and Programs in Biomedicine* **91**, 55–81 (July 2008).

Table 3: Detail comparison results in term of decimation and of presence of JTC correction for each type of events ($s_{plane} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $k = 0.4$).

					Decimation					
					1		2		3	
					JTC Tracking					
#	Movement	Speed	Amplitude	Direction	Classical	Histogram	Classical	Histogram	Classical	Histogram
1	None				57.75	83.58	57.39	83.31	54.22	79.95
2	Face rotation	Slow	Soft		89.07	89.07	88.14	88.14	74.34	79.94
3	Translation	Slow		Back	67.39	81.38	86.01	98.26	86.09	86.74
4	Translation	Slow		Front	63.91	72.74	80.78	100.00	78.44	88.94
5	Translation	Slow		Down	100.00	100.00	100.00	100.00	98.04	98.04
6	Translation	Slow		Up	100.00	100.00	87.74	87.74	81.13	81.13
7	Translation	Fast		Down	26.97	69.74	32.24	61.84	34.21	49.34
8	Translation	Fast		Up	20.23	56.30	29.03	48.09	31.67	46.04
9	Face rotation	Fast	Soft		61.51	91.51	36.51	87.09	30.70	72.91
10	Face rotation	Fast	High		19.68	46.96	16.53	51.42	16.63	37.12
11	Translation	Fast		Back	60.21	74.39	50.87	65.05	51.56	65.74
12	Translation	Fast		Front	57.86	78.93	44.34	79.56	49.06	77.67
13	Occlusion				18.85	41.97	19.44	44.18	30.93	40.21
14	Out of scope				28.66	39.10	35.22	62.69	60.90	61.19

[3] Chan, M., Campo, E., Estve, D., and Fourniols, J.-Y., "Smart homes-Current features and future perspectives," *Maturitas* **64**, 90–97 (October 2009).

[4] De Silva, L. C., Morikawa, C., and Petra, I. M., "State of the art of smart homes," *Engineering Applications of Artificial Intelligence* **25**(7), 1313–1321 (2012).

[5] Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., MacIntyre, B., Mynatt, E., Starner, T. E., and Newstetter, W., "The Aware Home: A Living Laboratory for Ubiquitous Computing Research," in [*Cooperative Buildings. Integrating Information, Organizations, and Architecture*], Streitz, N., Siegel, J., Hartkopf, V., and Konomi, S., eds., *Lecture Notes in Computer Science* **1670**, 191–198, Springer Berlin Heidelberg (1999).

[6] Cash, M., "Assistive technology and people with dementia," *Reviews in Clinical Gerontology* **13**, 313–319 (2003).

[7] Rantz, M., Marek, K., Aud, M., Tyrer, H., Skubic, M., Demiris, G., and Hussam, A., "A technology and nursing collaboration to help older adults age in place.," *Nursing outlook* **53**(1), 40–45 (2005).

[8] Demongeot, J., Virone, G., Duchene, F., Benchetrit, G., Herve, T., Noury, N., and Rialle, V., "Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people," *Comptes Rendus Biologies* **325**(6), 673–682 (2002).

[9] Williams, A., Ganesan, D., and Hanson, A., "Aging in place: fall detection and localization in a distributed smart camera network," in [*Proceedings of the 15th international conference on Multimedia*], *MULTIMEDIA '07*, 892–901, ACM, New York, NY, USA (2007).

[10] Tabar, A. M., Keshavarz, A., and Aghajan, H., "Smart home care network using sensor fusion and distributed vision-based reasoning," in [*Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*], *VSSN '06*, 145–154, ACM, New York, NY, USA (2006).

[11] Keshavarz, A., Tabar, A. M., and Aghajan, H., "Distributed vision-based reasoning for smart home care," in [*ACM SenSys Workshop on Distributed Smart Cameras (DSC06*], (2006).

[12] Aghajan, H., Augusto, J., Wu, C., McCullagh, P., and Walkden, J.-A., "Distributed Vision-Based Accident Management for Assisted Living," in [*Pervasive Computing for Quality of Life Enhancement*], Okadome, T., Yamazaki, T., and Makhtari, M., eds., *Lecture Notes in Computer Science* **4541**, 196–205, Springer Berlin Heidelberg (2007).

[13] Helal, S., Mann, W., El Zabadani, H., King, J., Kaddoura, Y., and Jansen, E., "The Gator Tech Smart House: a programmable pervasive space," *Computer* **38**, 50–60 (March 2005).

[14] Estudillo Valderrama, M., Roa, L., Reina Tosina, J., and Naranjo Hernandez, D., "Design and Implementation of a Distributed Fall Detection System-Personal Server," *Information Technology in Biomedicine, IEEE Transactions on* **13**, 874–881 (November 2009).

[15] De Silva, L. C. and Darussalam, B., "Audiovisual sensing of human movements for home-care and security in a smart environment," *Int. J. Smart Sens. Intell. Syst.* **1(1)**, 220–245 (2008).

[16] Chahuara, P., Portet, F., and Vacher, M., "Location of an inhabitant for domotic assistance through fusion of audio and non-visual data," in [*Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*], 242–245 (May 2011).

[17] Särelä, A., Korhonen, I., Lotjonen, J., Sola, M., and Myllymaki, M., "IST Vivago-an intelligent social and remote wellness monitoring system for the elderly," in [*Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*], 362–365 (April 2003).

[18] "Life Alert." `http://lifealert.com`.

[19] "Tunstall Telecare." `http://www.tunstallap.com`.

[20] Weaver, C. S. and Goodman, J. W., "A Technique for Optically Convolving Two Functions," *Appl. Opt.* **5**, 1248–1249 (July 1966).

[21] Shaou-Gang Miaou, Pei-Hsu Sung, and Chia-Yuan Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information," in [*Distributed Diagnosis and Home Healthcare, 2006. D2H2. 1st Transdisciplinary Conference on*], 39–42 (April 2006).

[22] Yilmaz, A., Javed, O., and Shah, M., "Object tracking: A survey," *ACM Comput. Surv.* **38** (December 2006).

[23] Lowe, D., "Distinctive image features from scale-invariant keypoints," *International journal of computer vision* **60**(2), 91–110 (2004).

[24] Jianbo Shi and Malik, J., "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**, 888–905 (August 2000).

[25] Caselles, V., Kimmel, R., and Sapiro, G., "Geodesic Active Contours," *International Journal of Computer Vision* **22**, 61–79 (1997).

[26] Stauffer, C. and Grimson, W., "Learning patterns of activity using real-time tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on DOI -10.1109/34.868677* **22**(8), 747–757 (2000).

[27] Papageorgiou, C., Oren, M., and Poggio, T., "A general framework for object detection," in [*Computer Vision, 1998. Sixth International Conference on*], 555–562 (January 1998).

[28] Alfalou, A. and Brosseau, C., "Understanding Correlation Techniques for Face Recognition: from basis to application," in [*Face Recognition*], Oravec, M., ed., 353–380, InTech (2010).

[29] Lugt, A. V., "Signal detection by complex spatial filtering," *IEEE Journals* **10**, 139–145 (1964).

[30] Katz, P., Alfalou, A., Brosseau, C., and Alam, M., "Correlation and Independent Component Analysis Based Approaches for Biometric Recognition," in [*Face Recognition: Methods, Applications and Technology*], Quaglia, A. and Epifano, C. M., eds., 201–229, NOVA Publisher (2012).

[31] Alfalou, A., Brosseau, C., Katz, P., and Alam, M., "Decision optimization for face recognition based on an alternate correlation plane quantification metric," *Opt. Lett.* **37**, 1562–1564 (May 2012).

[32] Ouerhani, Y., Jridi, M., Alfalou, A., and Brosseau, C., "Graphics Processor Unit Implementation of Correlation Technique using a Segmented Phase Only Composite Filter," in [*Optics Communications*], (289), 33–44 (2013).

[33] Bahram Javidi, "Nonlinear joint power spectrum based optical correlation," *Appl. Opt.* **28**, 2358–2367 (June 1989).

[34] Elbouz, M., Alfalou, A., and Brosseau, C., "Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance," *Optical Engineering* **50**(6), 067003 (2011).

[35] Viola, P. and Jones, M., "Rapid object detection using a boosted cascade of simple features," in [*Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*], **1**, I–511–I–518 vol.1 (2001).

[36] Lienhart, R. and Maydt, J., "An extended set of Haar-like features for rapid object detection," in [*Image Processing. 2002. Proceedings. 2002 International Conference on*], **1**, I–900–I–903 vol.1 (2002).

[37] Pele, O. and Werman, M., "The Quadratic-Chi Histogram Distance Family," in [*Computer Vision ECCV 2010*], Daniilidis, K., Maragos, P., and Paragios, N., eds., *Lecture Notes in Computer Science* **6312**, 749–762, Springer Berlin Heidelberg (2010).

[38] Comaniciu, D., Ramesh, V., and Meer, P., "Real-time tracking of non-rigid objects using mean shift," in [*Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*], **2**, 142–149 vol.2 (2000).

[39] Rowley, H., Baluja, S., and Kanade, T., "Rotation Invariant Neural Network-Based Face Detection," in [*Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*], 963 (June 1998).

Joint Transform Correlator pour le suivi de visages : application à la détection des chutes de la personne âgée

Philippe KATZ, Michael ARON, Ayman ALFALOU

VISION_L@BISEN, ISEN : Institut Supérieur de l'Électronique et du Numérique
20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France
philippe.katz@isen.fr, michael.aron@isen.fr, ayman.al-falou@isen.fr

Résumé – Dans un souci de maintien (sécurisé) à domicile des personnes dépendantes (e.g. personnes âgées), nous proposons une approche de détection des chutes basée sur l'étude de la trajectoire de la tête. Notre système est composé de différents modules (suivi, identification, fusion de données, décision...). Dans cet article, nous présentons une optimisation du suivi, basé sur un JTC (Joint Transform Correlator) et une comparaison d'histogrammes. Des expérimentations préliminaires de détection des chutes ont été réalisées, validant le principe.

Abstract – Seeking for a solution for dependent people (e.g. elderly) to remain safely at home, we propose a fall detection approach based on a head trajectory analysis. Our system is made up of various modules (tracking identification, data fusion, decision...). In this paper, we present an optimisation of the tracking stage based on a JTC (Joint Transform Correlator) and a histogram comparison. Preliminary experiments of fall detection have been performed, leading to the principle validation.

1 Introduction

Du fait du vieillissement de la population dans les pays industrialisés, les infrastructures traditionnelles (e.g. maisons de retraite) sont dans l'incapacité de couvrir le besoin croissant de prise en charge. Ainsi, l'urgence d'imaginer de nouvelles solutions s'est faite ressentir et le concept d'habitat intelligent a émergé. La chute d'une personne dépendante isolée étant le premier cas de décès accidentel, les solutions actuellement commercialisées utilisent en majorité des données issues de capteurs portés (bouton poussoir actionné par la personne elle-même, accéléromètres...). Cependant, le fait de porter le système engendre des contraintes : incapacité d'actionner l'alerte, casse du matériel lors de la chute... Afin de pallier ces problèmes, et en coopération avec Malakoff-Médérik, nous avons décidé d'utiliser une caméra vidéo pour la détection des chutes. Ainsi, nous proposons d'effectuer un suivi de la tête de la personne pour en extraire sa trajectoire et en déduire la chute éventuelle. Un grand nombre de méthodes de détection vidéo existent dans la littérature [1]. Deux familles majeures se distinguent : les méthodes numériques (e.g. détection de points d'intérêt [2], soustraction de fond [3]) et les méthodes optiques (e.g. Joint Transform Correlator [4]). Malgré un essoufflement de la production scientifique sur les corrélateurs, ceux-ci ont l'avantage de réaliser simultanément la détection, la localisation et l'identification de l'objet cible dans une scène [4].

Ainsi nous présentons dans cet article un système de suivi itératif basé sur le Joint Transform Correlator (JTC) et couplé avec une méthode numérique [5] pour proposer un système fiable. Nous introduisons tout d'abord la méthode de corrélation ainsi que notre algorithme de suivi par JTC. Puis nous présentons notre protocole d'expérimentation, permettant d'étudier l'impact des paramètres du JTC ainsi que les perfomances de suivi.

2 Détection des chutes par corrélation

2.1 Joint Transform Correlator : principe

Introduit par Weaver et Goodman [6], le JTC est une méthode de reconnaissance optique basée sur la comparaison d'une image cible (image à reconnaitre) et d'une image de référence dans un unique plan d'entrée. Son implémentation numérique s'effectue de la façon suivante : (i) un plan de Fourier est construit à partir de la transformée de Fourier (TF) du plan d'entrée ; (ii) le module au carré de ce plan est calculé ; (iii) la transformée de Fourier inverse (TF^{-1}) donne finalement le plan de corrélation. Ce plan contient trois pics principaux : (i) un pic central, ou « ordre zéro », correspondant à la somme des autocorrélations de chacune des images présentes dans le plan d'entrée ; (ii) deux pics périphériques, positionnés symétriquement par rapport au pic central, correspondant à la corrélation croisée entre les images cible et référence. L'intensité de ces pics de corrélation est conditionnée par le degré de similarité entre les images d'entrée. Quant à leur position, elle est dépendante de la localisation relative des images sur le plan d'entrée. Ce sont ces deux propriétés fondamentales qui permettent la détection et la localisation du visage de la personne dans l'image. Malheureusement, dans sa version classique, ce corrélateur a deux principaux défauts : l'ordre zéro, extrêmement

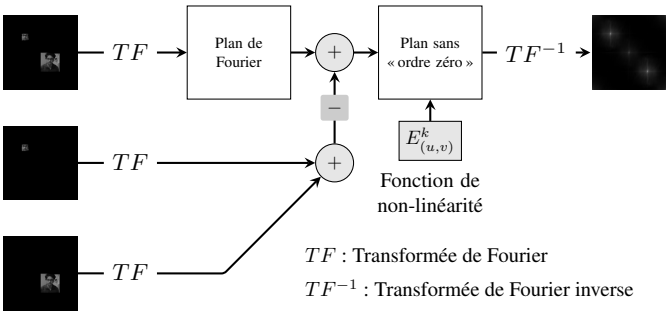


FIG. 1 – Joint Transform Correlator non-linéaire non-zéro.

intense par rapport aux pics de corrélation, et les pics de corrélation eux-mêmes, qui sont de faible intensité et très larges. La suppression de l'ordre zéro peut être effectuée en créant deux plans d'entrée supplémentaires, un pour chaque image contenue dans le plan d'entrée principal, et en déduisant leur spectre du plan de Fourier (Fig. 1).

Pour remédier au problème de faible intensité des pics de corrélation, Javidi et al. [7] proposent l'utilisation d'une fonction de non-linéarité dans le plan de Fourier, élevant ce dernier à une puissance $k < 1$. L'utilisation d'une valeur de k proche de 1 engendre de larges pics de corrélation de faible intensité, tandis qu'une valeur proche de 0 a l'effet inverse. Ainsi, le choix de la fonction de non-linéarité a un effet à la fois sur la robustesse du JTC et la précision de la localisation des pics. Il est donc nécessaire de déterminer expérimentalement un compromis entre ces deux propriétés. Cette étape de non-linéarité est appliquée après la suppression des spectres des images d'entrée (Fig. 1).

2.2 Application au suivi de visages

Notre algorithme est basé sur une méthode itérative : la région détectée dans l'image cible au temps t est utilisée comme image de référence au temps $t + 1$, la corrélation nous permettant d'effectuer dans un même processus la détection et le suivi. Le non-zero non-linear JTC (nz-nl-JTC) est initialisé par une image de référence, positionnée en haut à gauche du plan d'entrée. L'image cible est insérée sur le côté opposé. La position de l'objet dans l'image cible est obtenue à l'aide de la position relative des pics de corrélation. Cette position est finalement utilisée pour former l'image de référence pour l'itération suivante. Pour notre application de détection de chutes, nous proposons de choisir comme image de référence le visage de la personne. Notre méthode itérative est initialisée sur la première image de référence en utilisant un classifieur de Viola et Jones [8].

2.3 Optimisation du suivi

Malgré l'utilisation d'une fonction de non-linéarité, la faiblesse de l'intensité du pic de corrélation subsiste en cas de faible ressemblance de l'objet suivi entre les deux images (e.g. flou généré par une chute rapide). De plus, l'utilisation d'un suivi itératif rend définitive toute perte de l'objet suivi (e.g. une corrélation avec l'arrière-plan de l'image cible). Finalement, pour être en mesure d'effectuer le suivi en temps réel (i.e. entre deux captures d'images), il est nécessaire de réduire la taille du plan d'entrée. Pour pallier à cela, nous proposons deux optimisations : l'une basée sur une information déterministe (région d'intérêt), l'autre sur une comparaison d'histogrammes.

2.3.1 Définition d'une région d'intérêt

Nous introduisons une première optimisation consistant à limiter la recherche de l'objet suivi à une région donnée. Pour ce faire, nous définissons dans l'image cible une région d'intérêt autour de la position du visage au temps $t - 1$. Notre algorithme étant destiné à une application de détection de chutes, les mouvements ascendants sont par conséquent limités. Nous choisissons de conserver une plus vaste région sous le visage. La taille optimale de cette région, dépendante de la fréquence de rafraichissement de la caméra, a été déterminée expérimentalement (trois fois la région du visage, décalée d'un tiers vers le bas de l'image).

2.3.2 Mesure de la similarité des histogrammes

Comme explicité précédemment, leur incapacité à détecter une perte du suivi constitue le principal défaut des algorithmes itératifs. Pour remédier à cela, il est nécessaire d'introduire un critère supplémentaire. Ainsi, nous proposons l'utilisation de l'information contenue dans les histogrammes de l'image référence et de la région détectée. Un histogramme mesure la répartition de l'intensité des pixels : deux régions similaires visuellement auront donc des histogrammes semblables. Par conséquent, nous quantifions la similarité entre les images à l'aide du Chi Square de Pearson [9], défini par l'équation 1 :

$$X^2(H_{t-1}, H_t) = \sum_I \frac{(H_{t-1}(I) - H_t(I))^2}{H_{t-1}(I)} \qquad (1)$$

- H_{t-1} *and* H_t *correspondent aux histogrammes des régions détectées aux temps* t *et* $t - 1$*, respectivement,*
- I *correspond aux classes de l'histogramme.*

Concrètement, cette mesure de similarité permet de détecter une perte du suivi et de réinitialiser notre algorithme.

2.4 Application à la détection des chutes

Il est nécessaire de déterminer un critère caractérisant une chute. Une chute brutale peut-être définie par le passage dans un faible laps de temps d'une position debout à une position allongée. Ainsi, il est possible à l'aide de la vitesse verticale du visage de définir un seuil à partir duquel une chute est considérée comme ayant eu lieu. Malheureusement, ce type de critère ne prend pas en compte les chutes suivant un mouvement elliptique (de vitesse verticale plus faible). La formule décrite par l'équation 2 permet de prendre en compte à la fois un mouvement vertical et elliptique en pondérant le mouvement horizontal par un facteur $1/4$ (déterminé expérimentalement).

$$(y_t - y_{t-1}) + \frac{1}{4} \times |x_t - x_{t-1}| > Seuil \qquad (2)$$

– x_t *et* y_t *sont les coordonnées du visage dans l'image au temps t.*

Ce critère simple a été utilisé pour valider notre approche, mais nous prévoyons d'expérimenter d'autres critères dans de futurs travaux.

3 Expérimentations et résultats

Dans cette partie, nous décrivons tout d'abord les performances de notre système de suivi par rapport au JTC classique, ainsi que les apports de la correction par histogrammes (partie 3.2). L'application de cet algorithme à la détection des chutes est évalué en partie 3.3.

3.1 Protocole expérimental

Afin d'expérimenter notre algorithme, nous avons défini un protocole de tests et aménagé une salle de simulation (Fig. 2a) reproduisant une chambre d'hôpital ou de maison de retraite. Pour valider notre approche de détection des chutes, nous avons réalisé deux bases de données explorant les performances du suivi pour la première et de la détection de chutes pour la seconde.

Nous avons imaginé une large variété de scénarios composés de 14 différents événements (i.e. aucun mouvement, position debout, rotations du visage, translations, occlusions ou hors champ) de différentes vitesses ou amplitudes (Fig. 2b). La base de données, composée de 21087 images, a été enregistrée sur un total de 5 personnes de différentes origines ethniques. Finalement, l'emplacement du visage a été indiqué manuellement sur chaque image afin d'obtenir une « vérité terrain ».

(a) (b)

FIG. 2 – Protocole expérimental : (a) chambre de simulation ; (b) événements d'expérimentation (chutes et occlusions).

Une seconde base de données a été réalisée pour déterminer la capacité de détection des chutes du système. Elle est composée de 60 chutes – 20 chutes verticales (descente verticale du visage dans l'image) et 40 chutes elliptiques (dans les 2 directions). Quatre personnes ont participé à son enregistrement.

3.2 Performances du suivi JTC

Afin d'optimiser les performances du suivi par JTC, nous avons observé l'effet des différents paramètres disponibles dans notre environnement de test. Les meilleurs résultats sont obtenus avec : une décimation $d = 2$ (une image sur deux est prise en compte), un coefficient de non-linéarité $k = 0,4$, une taille de la région d'intérêt $s_{roi} = 3 \times s$ (s la région dans laquelle se situe le visage) et une taille de plan de corrélation $s_{corr} = (512 \times 512)$. Une étude détaillée est disponible dans [5].

Le tableau 1 présente une comparaison des résultats du suivi avec la« vérité terrain », obtenus en appliquant notre algorithme sur la base de 21 087 images (partie 3.1). Il présente les pourcentages d'images avec et sans suivi pour le suivi JTC classique ainsi que le nz-nl-JTC avec et sans correction par histogrammes (le critère de similarité 2.3.2 utilisé est $X^2 > 100$).

TAB. 1 – Comparaison entre le JTC classique et le nz-nl-JTC avec région d'intérêt (ROI) et avec et sans correction par histogrammes (Hist) – $s_{corr} = (512, 512)px$, $s_{roi} = (3 \times s)px$, $s = (78, 78)px$ and $k = 0.4$.

Suivi	JTC	nz-nl-JTC	
		ROI	ROI+Hist
Images avec suivi (%)	15, 35	58, 11	81, 46
Images sans suivi (%)	84, 65	41, 89	18, 54

On observe une amélioration significative du pourcentage d'images suivies avec l'utilisation du nl-nz-JTC par rapport au JTC classique, à savoir une augmentation de $42, 76pts$ du nombre d'images suivies. Malgré tout, le taux de $58, 11\%$ d'images suivies reste insuffisant pour notre application. Ce taux est amélioré de $23, 35pts$ par la correction par histogrammes, permettant d'obtenir un taux de $81, 46\%$ d'images suivies. Cependant, des pertes de suivi persistent, principalement lors de mouvements rapides ou d'occlusions [5], engendrant des limitations lors de l'étape suivante de détection des chutes.

3.3 Détection des chutes

Après avoir optimisé le suivi, nous observons les performances de notre critère de décision (décrit en 2.4). Le seuil est fixé expérimentalement à $450px$ sur un intervalle de $1, 34s$. Les résultats présentés sur le tableau 2 ont été réalisés sur l'ensemble de 60 chutes verticales et latérales (partie 3.1). Le nombre et le pourcentage de chutes détectées sont représentés sur les lignes 2 et 3. On observe un taux de $57, 5\%$ et de 60% de chutes elliptique et verticales reconnues, respectivement. Nous obtenons un total de $58, 34\%$ de chutes détectées.

Ce faible taux de détection est dû à une baisse des performances du suivi lors de mouvements rapides. De plus, l'absence d'information de profondeur permet une mesure de la vitesse uniquement à une distance fixe de la caméra.

TAB. 2 – Nombre et pourcentage de chutes (direction verticale, gauche, et droite) détectées par notre système (20 chutes simulées pour chaque situation).

Direction	Verticale	Gauche	Droite	TOTAL
Chutes détectées #	13	10	12	35
Chutes détectées %	65	50	60	58, 34
		57, 5		

4 Discussion

L'utilisation d'une mesure de similarité des histogrammes améliore significativement les performances du suivi par nz-nl-JTC, permettant d'atteindre un taux de $86, 41\%$ d'images suivies.

L'étape de détection de chutes correspond à une première expérimentation et nécessite des travaux plus étendus. Ses déficiences sont dues à la fois à l'étape de suivi, malheureusement insuffisante, et au critère de détection de la chute utilisé. En effet, différentes lacunes de notre algorithme persistent pour une telle application.

Premièrement, lors d'une perte de suivi au cours de la chute, l'étape d'initialisation du suivi ne permet pas d'obtenir un taux de détection du visage acceptable. De plus, des mouvements rapides engendrent un flou sur l'image capturée, perturbant la corrélation. Ces deux effets combinés augmentent à la fois la probabilité de décrochage du système tout en réduisant sa capacité à se réinitialiser.

Deuxièmement, le critère de détection n'est adapté qu'aux chutes brutales. Les chutes molles ou syncopales (le sujet se retient à un meuble ou perd connaissance) ne peuvent être détectées que par une méthode décision beaucoup plus développée, basée sur une analyse approfondie de la chute.

Finalement, des déplacements dans la profondeur influent à la fois sur le critère de détection et sur le JTC. La corrélation étant très sensible au changement d'échelle, la probabilité de perte de suivi en est d'autant plus augmentée. En outre, notre critère de détection ne peut prendre en compte des chutes dans cette direction.

Des améliorations peuvent être apportées. Nous développons un système multicaméra afin d'obtenir les informations de profondeur et de détecter la personne dans l'ensemble de la pièce. Également, notre méthode nécessite que le visage soit continuellement visible par la caméra [5]. Une détection du squelette améliorerait notre système, afin de connaitre la posture de la personne dans l'environnement.

5 Conclusion et perspectives

Cet article présente une étude de l'architecture JTC dans une application de détection de chutes de la personne âgée. Pour optimiser le suivi, deux améliorations de l'algorithme ont été apportées, à savoir la définition d'une région d'intérêt ainsi qu'un critère de similarité permettant une réinitialisation en cas d'échec du suivi.

En ce qui concerne notre méthode de suivi, bien que présentant des résultats prometteurs, elle doit être améliorée afin de prendre en compte la profondeur – une implantation multicaméra (stéréovision) est actuellement en cours. De plus, elle doit être en mesure d'offrir des informations de position et de posture de la personne lorsque le visage n'est plus dans le champ de vision de la caméra. Pour cela, nous proposons de fusionner notre algorithme avec des méthodes de détection de silhouette ou de squelette.

Notre critère de détection des chutes fait actuellement l'objet d'une étude plus approfondie, afin d'être à même de caractériser l'ensemble des différents types de chutes (chutes brutales, molles et syncopales).

Pour terminer, la réalisation d'une base de données étendue doit être réalisée pour expérimenter plus précisément notre algorithme ainsi que pour le comparer avec des méthodes existantes [10].

Remerciements Ces travaux sont supportés par le groupe Malakoff-Médérik et la société Open (Projet 0OR251).

Références

[1] A. Yilmaz, O. Javed and M. Shah, Object tracking : A survey, *ACM Comput. Surv.*, 38, 2006

[2] D. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision*, 60, 91–110, 2004

[3] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, Background modeling and subtraction of dynamic, *IEEE ICCV*, 1305–1312, 2003

[4] A. Alfalou and C. Brosseau, Understanding correlation techniques for face recognition : from basis to application, *Face Recognition*, M. Oravec (Ed.), 353–380, 2010

[5] P. Katz, M. Aron and A. Alfalou, Joint Transform Correlation for face tracking : elderly fall detection application, *SPIE 8748, Optical Pattern Recognition XXIV*, 87480I–14, 2013

[6] C. Weaver and J. W. Goodman, A technique for optically convolving two functions, *Appl. Opt.*, 5(7), 1248–1249, 1966

[7] B. Javidi, J. Wang and Q. Tang. Nonlinear joint transform correlators. *Pattern recognition*, 27(4), 523–542, 1994

[8] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE CVPR*, 1, 511–518, 2001

[9] O. Pele and M. Werman, The quadratic-chi histogram distance family, *Computer Vision ECCV*, 6312, 749–762, 2010

[10] C. Rougier, A. St-Arnaud, J. Rousseau and J. Meunier, Video surveillance for fall detection, *Video Surveillance*, Prof. Weiyao Lin (Ed.), InTech, 2011

SPIE Newsroom

10.1117/2.1201307.004994

A face-tracking system to detect falls in the elderly

Philippe Katz, Michael Aron, and Ayman Alfalou

An automated surveillance method that uses multiple image processing can detect, analyze, and track movements to identify emergency situations.

As life expectancy increases and birth rates fall, most industrialized countries anticipate a growing elderly population in the coming century. In Western Europe, for example, people aged over 60 represented 20% of the total population in 2000, but this number will reach 42% in 2050.[1] Given these projections, and the costs and logistics of caring for the elderly, it is generally recommended that the healthiest dependent people remain in their own homes, rather than transferring to an institutionalized setting. To realize this aim, care applications, or 'smart homes,' have evolved in recent decades.[2–12] These heterogeneous systems, designed to assist dependent people in everyday life, include automatic detection methods for falls—the primary cause of accidental death in isolated dependent people. Solutions currently available include wearable sensors (push-buttons[8] or accelerometers[5,8,9]), but these technologies have major drawbacks. For example, carelessness or cognitive trouble can lead to them being worn intermittently, and the wearer of the sensor needs to be conscious to press the button. Furthermore, when a loss of consciousness occurs slowly, it is undetectable by this kind of technology.

Consequently, we require a system that is able to interpret a situation and detect and analyze a movement. We propose an automated and stand-alone surveillance method, fully integrated within the environment (see Figure 1). A large number of sensors set up in the home would collect different kinds of accessible data: audio, video, IR, or pressure (from sensors embedded in furniture). Information from these would pass to a local calculation unit for testing and analysis. Thus, it would be possible to consider a large variety of situations such as falls, unusual inaction, or a sudden change in habits. Information about these events would go to emergency services, and would provide diagnostic information to health practitioners. Furthermore, an alert would go to relatives by Short Message Service or email.

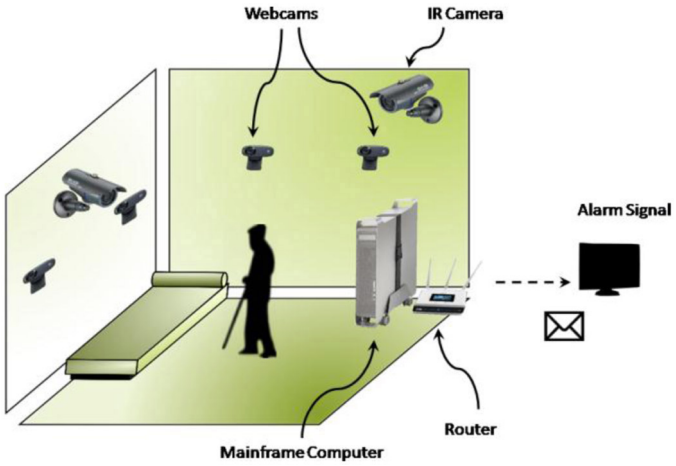


Figure 1. *An overview of the fall detection system.*

Based on the concept of a fall as a transition from standing to lying down, we tracked the position of a subject's face to assess temporal and spatial information. At present, our work focuses on this tracking stage.[13] Our system has the advantages of being relatively simple and able to simultaneously process detection, identification, and localization. A Fourier transform is applied on an entry plane composed of a reference and a target image (the face to be recognized), and an inverse Fourier transform yields a correlation plane. In this study we use a joint transform correlator (JTC),[14] an image processing technique that can be used to compare several images in parallel, and which is particularly suitable for tracking situations.

The correlation plane given by a JTC implementation contains two cross-correlation peaks whose location depends on the relative position of reference and target images in the entry plane. This allows the localization of a target motif (namely a face) in a scene. An iterative algorithm—in which the reference image at each timestamp (time t) is replaced by the previously detected face in the target image ($t-1$)—makes face tracking possible in each video frame, taking into account the different variations of

Continued on next page

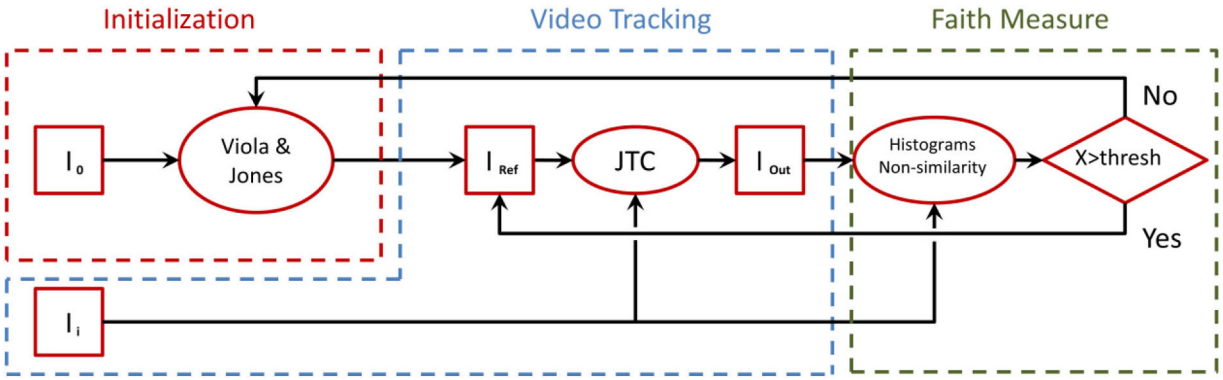


Figure 2. *The joint transform correlator (JTC) algorithm with a synopsis of the histogram optimization. I_0 refers to the initialization image. I_i is the i^{th} image of the video sequence under consideration. I_{Ref} and I_{Out} are the reference and output images, respectively, and X denotes the result of our histogram similarity criterion, which is compared with a given threshold (thresh).*

our tracked motif in time (see Figure 2). The algorithm initialization ($t = 0$) is performed by means of a Viola-Jones object detection framework.[15] To avoid false detection cases (correlation with the scene background, for example), we realize a histogram comparison between the target and reference image (the person's face). We can detect a large inter-frame variation, making it possible to re-initialize our algorithm if there is a loss of tracking.

We produced an experimental setup to test the reliability of our approach. First, we created a reproduction of a hospital or retirement home (see Figure 3). Second, we imagined a wide variety of scenarios (where the subject is facing away from the detector, rotates, or falls, or where the face is hidden by another object) comprising 21,087 frames (see Figure 4). We recorded the face position on each frame manually, leading to 'ground truth,' where in each frame we manually localize the position of the head and register it, so that the information can be used to evaluate the algorithm. Table 1 presents a comparison between the iterative JTC algorithm with and without the histogram similarity stage. The effect of histogram correction is noticeable, giving an improvement of 23 percentage points.

Figure 3. *The simulation room, set up as a hospital or retirement home for our experiment.*

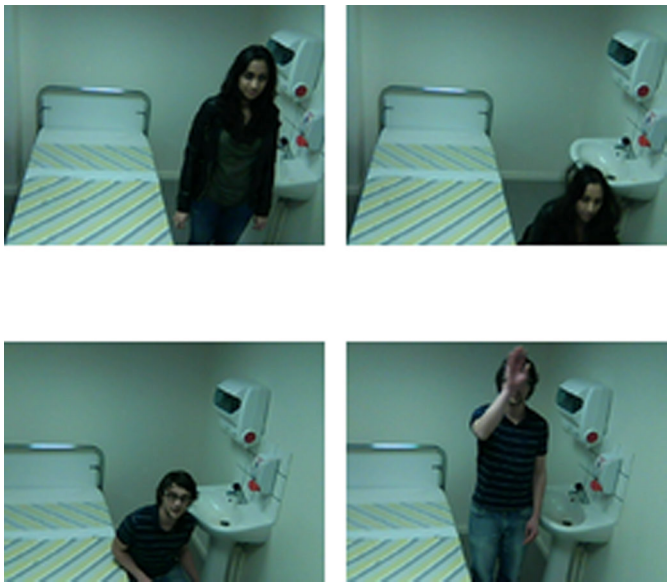


Figure 4. *Simulated falls and occlusions used to test the algorithm.*

Finally, we experimented with fall detection using a naive method based on speed measurement. A fall is detected when the downward vertical speed of a face across successive video

Continued on next page

Table 1. *Analysis of 21,087 frames using a joint transform correlator algorithm, with and without histogram correction. The first line of data (tracked pictures) presents the percentage of correctly detected faces.*

Tracking	Without histogram	Histogram
Tracked pictures (%)	58.11	81.46
Non-tracked pictures (%)	41.89	18.54

Table 2. *Number and percentage of falls detected by our algorithm for vertical, left, and right directions (20 falls simulated for each situation).*

	Vertical	Left	Right	Total
Recognized falls	13	10	12	35
Recognized falls (%)	65	50	60	58.34

frames exceeds a certain threshold. We also considered the horizontal speed for elliptical falls, weighting it by a 1/4 factor, yielding the formula

$$(y_t - y_{t-1}) + \frac{1}{4}|x_t - x_{t-1}| > Threshold,$$

where x_t and y_t are the face coordinates at time t.

The results, obtained on a set of 60 falls (vertical, left, and right), are shown in Table 2. We correctly detected 58% of the total falls. Various factors affected the result. The speed measured depends on the distance between the subject and the camera. If the face is obscured during a fall, the Viola-Jones detector may not be able to re-initialize the algorithm, and the naive fall detection method is unsuitable for slow falls and for when the face follows an elliptical trajectory.

Our method can simultaneously detect, localize, and identify the person. Furthermore, it can accurately perform a tracking process. Unfortunately, that process still suffers from some limitations, and the correlation has to be considered as a baseline method, to be improved in future work. A background subtraction (where we define the background scene with a fixed camera, and eliminate it from the results) may be an appropriate enhancement of our system, as would silhouette and skeleton detections for posture identification, which could be fuzzed with our system. Finally, we need to compare our technique with other fall detection systems, using an extended experimental database.[16]

This work is supported by Project 0OR251, a collaboration between the Malakoff-Médérik Group, Open Society, and ISEN-Brest.

Author Information

Philippe Katz, Michael Aron, and Ayman Alfalou
Vision Laboratory
Institut Supérieur d'Électronique et du Numérique (ISEN)
Brest, France

Philippe Katz received his engineering diploma from ISEN-Brest and his MSc in signals and images in biology and medicine from the University of Brest in 2011. Since then, he has been a PhD student at ISEN. His research interests include image and signal processing and smart homes.

Michael Aron received an engineering diploma in 2002 from Polytech-Sophia, University of Nice-Sophia Antipolis. He received his PhD in computer science from the University of Lorraine in 2009, and conducted his image processing post-doctoral research at The French Research Institute for Exploitation of the Sea. Since 2011, he has been an associate professor at ISEN-Brest. His research topics include computer vision and image processing.

Ayman Alfalou's research interests are in optical engineering, optical information processing, signal and image processing, telecommunications, and optoelectronics. He has published more than 110 refereed journal articles or conference papers, and is a senior member of SPIE, the Optical Society of America, and the Institute of Electrical and Electronics Engineers, and is a member of the Institute of Physics.

References

1. W. Lutz, W. Sanderson, and S. Scherbov, *The coming acceleration of global population ageing*, **Nature 451**, pp. 716–719, 2008.
2. M. Chan, D. Estve, C. Escriba, and E. Campo, *A review of smart homes: present state and future challenges*, **Computer Methods Programs Biomed. 91**, pp. 55–81, 2008.
3. L. C. De Silva and B. Darussalam, *Audiovisual sensing of human movements for home-care and security in a smart environment*, **Int'l J. Smart Sensing Intell. Syst. 1**, pp. 220–245, 2008.
4. J. Demongeot, G. Virone, F. Duchene, G. Benchetrit, T. Herve, N. Noury, and V. Rialle, *Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people*, **Comptes Rendus Biologies 325** (6), pp. 673–682, 2002.
5. A. Keshavarz, A. M. Tabar, and H. Aghajan, *Distributed vision-based reasoning for smart home care*, **ACM Sensys Workshop Distributed Smart Cameras**, 2006.
6. http://lifealert.com Life Alert: a medical alert system for home health emergencies. Accessed 4 July 2013.
7. S. G. Miaou, P. H. Sung, and C. Y. Huang, *A customized human fall detection system using omni-camera images and personal information*, **Proc. Transdisciplinary Conf. Distributed Diagnosis Home Healthcare D2H2**, pp. 39–42, 2006.
8. A. Särelä, I. Korhonen, J. Lotjonen, M. Sola, and M. Myllymaki, *IST Vivago–an intelligent social and remote wellness monitoring system for the elderly*, **IEEE EMBS Special Topic Conf. Inf. Technol. Appl. Biomed.**, pp. 362–365, 2003.
9. A. M. Tabar, A. Keshavarz, and H. Aghajan, *Smart home care network using sensor fusion and distributed vision-based reasoning*, **ACM Int'l Workshop Video Surveillance Sensor Networks**, pp. 145–154, 2006.

Continued on next page

10. http://www.tunstallap.com Tunstall Healthcare, a provider of technology and services to people with long-term health and care needs. Accessed 4 July 2013.
11. A. Williams, D. Ganesan, and A. Hanson, *Aging in place: fall detection and localization in a distributed smart camera network*, **ACM Int'l Conf. Multimedia**, pp. 892–901, 2007.
12. G. Demiris and B. K. Hensel, *Technologies for an aging society: a systematic review of "smart home" applications*, **Yearbook Med. Inf. 31**, pp. 33–40, 2008.
13. P. Katz, M. Aron, and A. Alfalou, *Joint transform correlation for face tracking: elderly fall detection application*, **Proc. SPIE 8748**, p. 87480I, 2013. doi:10.1117/12.2016413
14. C. S. Weaver and J. W. Goodman, *A technique for optically convolving two functions*, **Appl. Opt. 5**, pp. 1248–1249, 1966.
15. P. Viola and M. Jones, *Rapid object detection using a boosted cascade of simple features*, **IEEE Computer Soc. Conf. Computer Vision Pattern Recognition 1**, pp. 511–518, 2001.
16. C. Rougier, A. St-Arnaud, J. Rousseau, and J. Meunier, *Video surveillance for fall detection*, **Int'l Conf. Innovative Technol.**, pp. 357–382, 2011.

Annexes

Annexe A

Base d’apprentissage

FIGURE A.1 – Personne 1

FIGURE A.2 – Personne 2

FIGURE A.3 – Personne 3

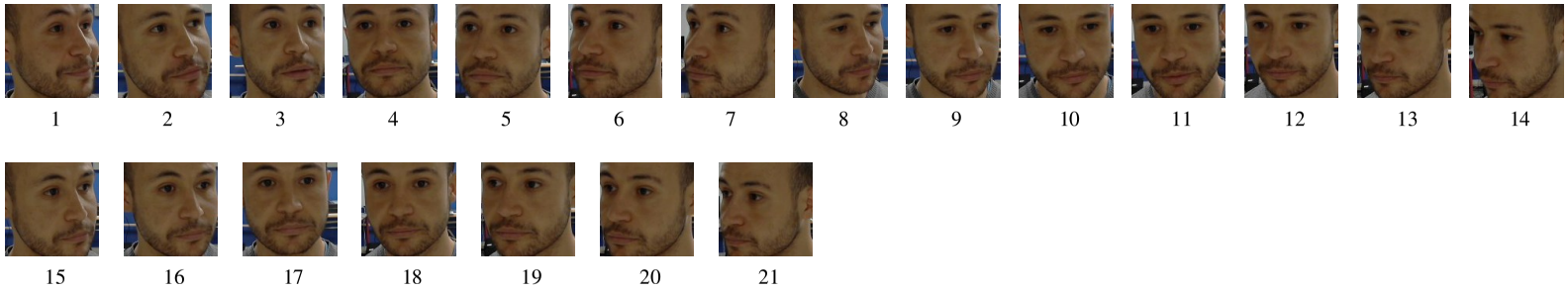


FIGURE A.4 – Personne 4

FIGURE A.5 – Personne 5

FIGURE A.6 – Personne 6

FIGURE A.7 – Personne 7

FIGURE A.8 – Personne 8

Annexe B

Base de test

FIGURE B.1 – Personne 1

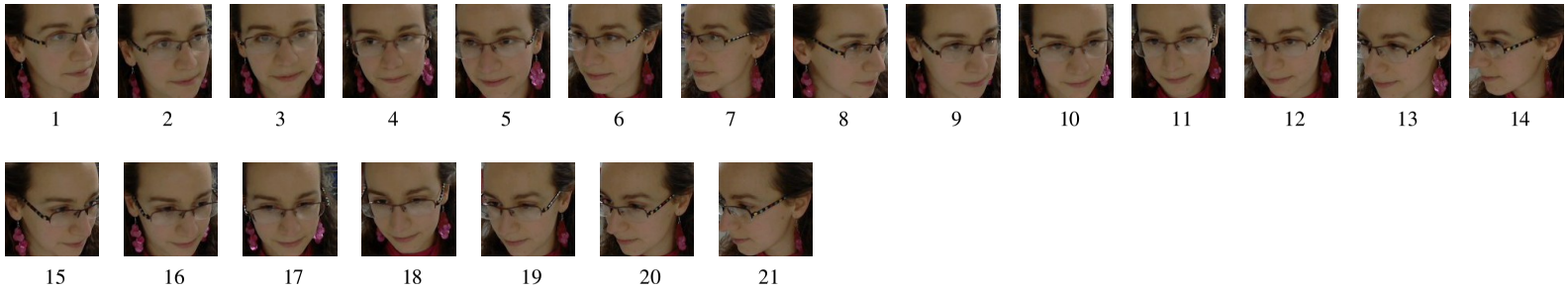


FIGURE B.2 – Personne 2

Bibliographie

Bibliographie

[1] G. Demiris and B. K. Hensel, "Technologies for an aging society : a systematic review of "smart home" applications.," *Yearbook of Medical Information*, vol. 31, no. 943-4747, pp. 33–40, 2008.

[2] W. Lutz, W. Sanderson, and Scherbov, "The coming acceleration of global population ageing,," *Nature*, vol. 451, pp. 716–719, 2008.

[3] N. Blanpain and O. Chardon, "Projections de population à l'horizon 2060. un tiers de la population âgé de plus de 60 ans," *Insee première*, vol. 1320, 2010.

[4] J. Charpin, A. Caillaud, and A. Dugué, *Les personnes âgées*. INSEE, 2005.

[5] C. Goillot and P. Mormiche, "Les enquêtes handicaps-incapacités-dépendances de 1998 et 1999 : résultats détaillés.," *INSEE : Paris*, p. 229, 2003.

[6] M. Chan, D. Esteve, C. Escriba, and E. Campo, "A review of smart homes : Present state and future challenges," *Computer Methods and Programs in Biomedicine*, vol. 91, no. 1, pp. 55–81, 2008.

[7] M. Chan, E. Campo, D. Esteve, and J.-Y. Fourniols, "Smart homes-Current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–97, 2009.

[8] L. C. De Silva, C. Morikawa, and I. M. Petra, "State of the art of smart homes," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1313–1321, 2012.

[9] P. Dargent-Molina and G. Breart, "Epidémiologie des chutes et des traumatismes liés aux chutes chez les personnes âgées," *Revue d'épidémiologie et de santé publique*, vol. 43, no. 1, pp. 72–83, 1995.

[10] L. Rubenstein, "Falls in older people : epidemiology, risk factors and strategies for prevention.," *Age Ageing*, vol. 35, no. S2, pp. ii37–ii41, 2006.

[11] J. Perry, S. Kellog, S. Vaidya, J.-H. Youn, H. Ali, and H. Sharif, "Survey and evaluation of real-time fall detection approaches," in *High-Capacity Optical Networks and Enabling Technologies (HONET), 2009 6th International Symposium*, pp. 158–164, 2009.

[12] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, and J. Lundy, "Fall detection - principles and methods," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 1663–1666, 2007.

[13] X. Yu, "Approaches and principles of fall detection for elderly and patient," in *e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference*, pp. 42–47, 2008.

[14] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection : Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, 2013. Special issue : Behaviours in video.

[15] N. Noury, P. Rumeau, A. Bourke, G. Òlaighin, and J. Lundy, "A proposal for the classification and evaluation of fall detectors," *IRBM*, vol. 29, no. 6, pp. 340–349, 2008.

[16] C. Rougier, *Vidéosurveillance intelligente pour la détection de chutes chez les personnes âgées*. PhD thesis, Université de Montréal, 2010.

[17] G. Wu, "Distinguishing fall activities from normal activities by velocity characteristics," *Journal of Biomechanics*, vol. 33, no. 11, pp. 1497–1500, 2000.

[18] D. Ding, R. A. Cooper, P. F. Pasquina, and L. Fici Pasquina, "Sensor technology for smart homes," *Maturitas*, vol. 69, no. 2, pp. 131–136, 2011.

[19] "Life Alert." `http://lifealert.com`, 2012.

[20] A. Särelä, I. Korhonen, J. Lotjonen, M. Sola, and M. Myllymaki, "IST Vivago-an intelligent social and remote wellness monitoring system for the elderly," in *IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 362–365, 2003.

[21] A. M. Tabar, A. Keshavarz, and H. Aghajan, "Smart home care network using sensor fusion and distributed vision-based reasoning," in *ACM international workshop on Video surveillance and sensor networks*, pp. 145–154, ACM, 2006.

[22] A. Keshavarz, A. M. Tabar, and H. Aghajan, "Distributed vision-based reasoning for smart home care," in *ACM SenSys Workshop on Distributed Smart Cameras DSC*, 2006.

[23] J. Demongeot, G. Virone, F. Duchene, G. Benchetrit, T. Herve, N. Noury, and V. Rialle, "Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people," *Comptes Rendus Biologies*, vol. 325, no. 6, pp. 673–682, 2002.

[24] M. Estudillo Valderrama, L. Roa, J. Reina Tosina, and D. Naranjo Hernandez, "Design and Implementation of a Distributed Fall Detection System-Personal Server," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 874–881, 2009.

[25] M. Nyan, F. Tay, A. Tan, and K. Seah, "Distinguishing fall activities from normal activities by angular rate characteristics and high-speed camera characterization," *Medical Engineering and Physics*, vol. 28, no. 8, pp. 842–849, 2006.

[26] A. Bourke and G. Lyons, "A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor," *Medical Engineering & Physics*, vol. 30, no. 1, pp. 84 – 90, 2008.

[27] S. Fleck and W. Strasser, "Smart camera based monitoring system and its application to assisted living," *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1698–1714, 2008.

[28] "Tunstall Telecare." `http://www.tunstallap.com`, 2012.

[29] T. Zhang, J. Wang, P. Liu, and J. Hou, "Fall detection by embedding an accelerometer in cellphone and using kfd algorithm," *International Journal of Computer Science and Network Security*, vol. 6, no. 10, pp. 277–284, 2006.

[30] J. Dai, X. Bai, Z. Yang, Z. Shen, and D. Xuan, "Mobile phone-based pervasive fall detection," *Personal Ubiquitous Comput.*, vol. 14, no. 7, pp. 633–643, 2010.

[31] L. Ni, D. Zhang, and M. Souryal, "RFID-based Localization and Tracking Technologies," *Wireless Communications, IEEE DOI-10.1109/MWC.2011.5751295*, vol. 18, no. 2, pp. 45–51, 2011.

[32] P. A. Cavalcante Aguilar, J. Boudy, D. Istrate, B. Dorizzi, J.-L. Baldinger, T. Guettari, I. Belfeki, and H. Medjahed, "Fusion multi-capteurs hétérogène basée sur un réseau d'évidence pour la détection de chute," in *ASSISTH '11 : 2e Conférence Internationale sur l'Accessibilité et les Systèmes de Suppléance aux personnes en sItuaTions de Handicap*, 2011.

[33] P. A. Cavalcante Aguilar, *Réseaux Évidentiels pour la fusion de données multimodales hétérogènes : application à la détection de chutes*. PhD thesis, Télécom Paris Sud, 2012.

[34] S. Helal, W. Mann, H. El Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The Gator Tech Smart House : a programmable pervasive space," *Computer*, vol. 38, no. 3, pp. 50–60, 2005.

[35] P. Srinivasan, D. Birchfield, G. Qian, and A. Kidaneé, "Design of a pressure sensitive floor for multimodal sensing," in *Information Visualisation, 2005. Proceedings. Ninth International Conference*, pp. 41–46, 2005.

[36] Y. Nishida, T. Hori, T. Suehiro, and S. Hirai, "Sensorized environment for self-communication based on observation of daily human behavior," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1364–1372, 2000.

[37] M. Alwan, P. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder, "A smart and passive floor-vibration based fall detector for elderly," in *Information and Communication Technologies, 2006. ICTTA '06. 2nd*, vol. 1, pp. 1003–1007, 2006.

[38] Y. Zigel, D. Litvak, and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound – proof of concept on human mimicking doll falls," *Biomedical Engineering, IEEE Transactions*, vol. 56, no. 12, pp. 2858–2867, 2009.

[39] A. Williams, D. Ganesan, and A. Hanson, "Aging in place : fall detection and localization in a distributed smart camera network," in *International conference on Multimedia*, pp. 892–901, ACM, 2007.

[40] H. Aghajan, J. Augusto, C. Wu, P. McCullagh, and J.-A. Walkden, "Distributed Vision-Based Accident Management for Assisted Living," in *Pervasive Computing for Quality of Life Enhancement* (T. Okadome, T. Yamazaki, and M. Makhtari, eds.), vol. 4541 of *Lecture Notes in Computer Science*, pp. 196–205, Springer Berlin Heidelberg, 2007.

[41] L. C. De Silva and B. Darussalam, "Audiovisual sensing of human movements for home-care and security in a smart environment," *Int. J. Smart Sens. Intell. Syst.*, vol. 1, no. 1, pp. 220–245, 2008.

[42] "Edao." `http://www.edao.com/etablissement/comment-ca-marche/`, 2014.

[43] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier, "Fall detection from depth map video sequences," in *Toward Useful Services for Elderly and People with Disabilities* (B. Abdulrazak, S. Giroux, B. Bouchard, H. Pigot, and M. Mokhtari, eds.), vol. 6719 of *Lecture Notes in Computer Science*, pp. 121–128, Springer Berlin Heidelberg, 2011.

[44] "Kinect, 3d-sensing technology." `http://www.primesense.com`, 2014.

[45] P. Chahuara, F. Portet, and M. Vacher, "Location of an inhabitant for domotic assistance through fusion of audio and non-visual data," in *International Conference on Pervasive Computing Technologies for Healthcare PervasiveHealth*, pp. 242–245, 2011.

[46] E. Auvinet, L. Reveret, A. St-Arnaud, J. Rousseau, and J. Meunier, "Fall detection using multiple cameras," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 2554–2557, 2008.

[47] S. McKenna and H. Charif, "Summarising contextual activity and detecting unusual inactivity in a supportive home environment," *Pattern Analysis and Applications*, vol. 7, no. 4, pp. 386–401, 2004.

[48] B. Jansen and R. Deklerck, "Context aware inactivity recognition for visual fall detection," in *Pervasive Health Conference and Workshops, 2006*, pp. 1–4, 2006.

[49] J. Tao, M. Turjo, M. Wong, M. Wang, and Y. Tan, "Fall incidents detection for intelligent video surveillance," in *Information, Communications and Signal Processing, 2005 Fifth International Conference*, pp. 1590–1594, 2005.

[50] Shaou-Gang Miaou, Pei-Hsu Sung, and Chia-Yuan Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information," in *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare D2H2*, pp. 39–42, 2006.

[51] C.-W. Lin and Z.-H. Ling, "Automatic fall incident detection in compressed video for intelligent homecare," in *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference*, pp. 1172–1177, 2007.

[52] H. Foroughi, B. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference*, pp. 219–224, 2008.

[53] H. Foroughi, A. Rezvanian, and A. Paziraee, "Robust fall detection using human shape and multi-class support vector machine," in *Computer Vision, Graphics Image Processing, 2008. ICVGIP '08. Sixth Indian Conference*, pp. 413–420, 2008.

[54] H. Foroughi, A. Naseri, A. Saberi, and H. Yazdi, "An eigenspace-based approach for human fall detection using integrated time motion image and neural network," in *Signal Processing, 2008. ICSP 2008. 9th International Conference*, pp. 1499–1503, 2008.

[55] R. Cucchiara, A. Prati, and R. Vezzani, "A multi-camera vision system for fall detection and alarm generation," *Expert Systems*, vol. 24, no. 5, pp. 334–345, 2007.

[56] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," *Computer Vision and Image Understanding*, no. 1, pp. 80–89, 2009.

[57] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *Circuits and Systems for Video Technology, IEEE Transactions*, vol. 21, no. 5, pp. 611–622, 2011.

[58] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic posture classification for human-behavior analysis," *Systems, Man and Cybernetics, Part A : Systems and Humans, IEEE Transactions*, vol. 35, no. 1, pp. 42–54, 2005.

[59] N. Thome and S. Miguet, "A HHMM-based approach for robust fall detection," in *9th International Conference on Control, Automation, Robotics & Vision (IEEE ICARCV)*, pp. 1–8, IEEE CNF, 2006.

[60] C. Rougier and J. Meunier, "Demo : Fall detection using 3d head trajectory extracted from a single camera video sequence," *Journal of Telemedicine and Telecare*, vol. 11, no. 4, 2005.

[61] D. Dementhon and L. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, no. 1-2, pp. 123–141, 1995.

[62] C. Rougier, A. St-Arnaud, J. Rousseau, and J. Meunier, "Video surveillance for fall detection," in *Int'l Conf. Innovative Technol.*, pp. 978–953., 2011.

[63] L. Hazelhoff, J. A Han, and P. de With, *Advanced Concepts for Intelligent Vision Systems*, ch. Video-Based Fall Detection in the Home Using Principal Component Analysis, pp. 298–309. Springer Berlin Heidelberg, 2008.

[64] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Fall detection with multiple cameras : An occlusion-resistant method based on 3-d silhouette vertical distribution," *Information Technology in Biomedicine, IEEE Transactions*, vol. 15, no. 2, pp. 290–300, 2011.

[65] A. Yilmaz, O. Javed, and M. Shah, "Object tracking : A survey," *ACM Comput. Surv.*, vol. 38, 2006.

[66] H. P. Moravec, "Visual mapping by a robot rover," in *Proceedings of the 6th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'79, pp. 598–600, Morgan Kaufmann Publishers Inc., 1979.

[67] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15, p. 50, 1988.

[68] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, vol. 81, pp. 674–679, 1981.

[69] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference*, vol. 2, pp. II–257–II–263, 2003.

[70] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[71] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder : real-time tracking of the human body," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 19, no. 7, pp. 780–785, 1997.

[72] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[73] N. Oliver, B. Rosario, and A. Pentland, "A Bayesian computer vision system for modeling human interactions," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 22, no. 8, pp. 831–843, 2000.

[74] D. Comaniciu and P. Meer, "Mean shift : a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 24, no. 5, pp. 603–619, 2002.

[75] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Computer Vision, 1995. Proceedings., Fifth International Conference*, pp. 694–699, 1995.

[76] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*, pp. 23–37, Springer, 1995.

[77] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, ACM, 1992.

[78] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human Activity Detection from RGBD Images," *CoRR*, vol. 1107.0169, 2011.

[79] A. Alfalou, C. Brosseau, and M. S. Alam, "Smart pattern recognition," in *SPIE Optical Pattern Recognition*, vol. 8748, pp. 874809–874809–23, SPIE, 2013.

[80] M. Elbouz, F. Bouzidi, A. Alfalou, C. Brosseau, I. Leonard, and B.-E. Benkelfat, "Adapted all-numerical correlator for face recognition applications," in *SPIE Optical Pattern Recognition*, vol. 8748, pp. 874807–874807–8, SPIE, 2013.

[81] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[82] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Analysis & Applications*, vol. 9, no. 2, pp. 273–292, 2006.

[83] I. T. Joliffe, "Principal Component Analysis," *New York : Springer-Verlag*, 1986.

[84] P. Comon, "Independent component analysis, A new concept ?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[85] A. Alfalou and C. Brosseau, "Robust and discriminating method for face recognition based on correlation technique and independent component analysis model," *Opt. Lett.*, vol. 36, no. 5, pp. 645–647, 2011.

[86] A. Alfalou and C. Brosseau, "Understanding Correlation Techniques for Face Recognition : from basis to application," in *Face Recognition* (M. Oravec, ed.), pp. 353–380, InTech, 2010.

[87] P. Katz, A. Alfalou, C. Brosseau, and M. Alam, "Correlation and Independent Component Analysis Based Approaches for Biometric Recognition," in *Face Recognition : Methods, Applications and Technology* (A. Quaglia and C. M. Epifano, eds.), pp. 201–229, NOVA Publisher, 2012.

[88] A. Alfalou, C. Brosseau, P. Katz, and M. Alam, "Decision optimization for face recognition based on an alternate correlation plane quantification metric," *Opt. Lett.*, vol. 37, no. 9, pp. 1562–1564, 2012.

[89] Y. Ouerhani, M. Jridi, and A. AlFalou, "Fast face recognition approach using a graphical processing unit "GPU"," pp. 80–84, 2010.

[90] M. Plancherel and M. Leffler, "Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies," *Rendiconti del Circolo Matematico di Palermo*, vol. 30, no. 1, pp. 289–335, 1910.

[91] Y. Ouerhani, M. Jridi, A. Alfalou, and C. Brosseau, "Graphics Processor Unit Implementation of Correlation Technique using a Segmented Phase Only Composite Filter," in *Optics Communications*, no. 289, pp. 33–44, 2013.

[92] A. Alfalou and A. Mansour, "Double random phase encryption scheme to multiplex and simultaneous encode multiple images," *Appl. Opt.*, vol. 48, no. 31, pp. 5933–5947, 2009.

[93] A. A. S. Awwal, M. A. Karim, and S. R. Jahan, "Improved correlation discrimination using an amplitude-modulated phase-only filter," *Applied Optics*, vol. 29, no. 2, pp. 233–236, 1990.

[94] D. Casasent and G. Ravichandran, "Advanced distortion-invariant minimum average correlation energy (mace) filters," *Applied Optics*, vol. 31, no. 8, pp. 1109–1116, 1992.

[95] A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Journals*, vol. 10, pp. 139–145, 1964.

[96] A. Mahalanobis, B. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Applied Optics*, vol. 26, no. 17, pp. 3633–3640, 1987.

[97] A. A. S. Awwal, "What can we learn from the shape of a correlation peak for position estimation ?," *Applied optics*, vol. 49, no. 10, pp. B40–B50, 2010.

[98] I. Leonard, A. Alfalou, and C. Brosseau, "Spectral optimized asymmetric segmented phase-only correlation filter," *Applied optics*, vol. 51, no. 14, pp. 2638–2650, 2012.

[99] F. M. Dickey and L. A. Romero, "Dual optimality of the phase-only filter," *Opt. Lett*, vol. 14, pp. 4–5, 1989.

[100] J. Horner, "Metrics for assessing pattern-recognition performance," *Appl. Opt.*, vol. 31, no. 2, pp. 165–166, 1992.

[101] B. V. K. V. Kumar and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.*, vol. 29, no. 20, pp. 2997–3006, 1990.

[102] Michel Barret, *Traitement Statistique du signal*. 2009.

[103] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[104] Y. Ouerhani, *Contribution à la définition, à l'optimisation et à l'implantation d'IP de traitement du signal et des données en temps réel sur des cibles programmables*. PhD thesis, Université de Bretagne occidentale-Brest, 2012.

[105] A. Oppenheim and J. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[106] J. L. Horner, B. Javidi, and J. Wang, "Analysis of the binary phase-only filter," *Optics communications*, vol. 91, no. 3, pp. 189–192, 1992.

[107] B. Kumar, "Partial information filters," *Digital signal processing*, vol. 4, no. 3, pp. 147–153, 1994.

[108] J. Ding, M. Itoh, and T. Yatagai, "Design of optimal phase-only filters by direct iterative search," *Optics communications*, vol. 118, no. 1, pp. 90–101, 1995.

[109] B. V. Kumar, F. M. Dickey, J. M. Connelly, and L. A. Romero, "Complex ternary matched filters yielding high signal-to-noise ratios," *Optical engineering*, vol. 29, no. 9, pp. 994–1001, 1990.

[110] M. A. Kaura and W. T. Rhodes, "Optical correlator performance using a phase-with-constrained-magnitude complex spatial filter," *Applied optics*, vol. 29, no. 17, pp. 2587–2593, 1990.

[111] R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," *Applied optics*, vol. 28, no. 22, pp. 4865–4869, 1989.

[112] F. M. Dickey and B. D. Hansche, "Quad-phase correlation filters for pattern recognition," *Applied optics*, vol. 28, no. 9, pp. 1611–1613, 1989.

[113] B. D. Hansche, J. J. Mason, and F. M. Dickey, "Quad-phase-only filter implementation," *Applied optics*, vol. 28, no. 22, pp. 4840–4844, 1989.

[114] P. Réfrégier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and horner efficiency," *Optics Letters*, vol. 16, no. 11, pp. 829–831, 1991.

[115] A. Alfalou, G. Keryer, J.-L. de Bougrenet de La Tocnaye, *et al.*, "Optical implementation of segmented composite filtering," *Applied optics*, vol. 38, no. 29, pp. 6129–6135, 1999.

[116] N. Gourier, D. Hall, and J. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *Pointing ICPR on International Workshop on Visual Observation of Deictic Gestures*, 2004.

[117] C. S. Weaver and J. W. Goodman, "A Technique for Optically Convolving Two Functions," *Appl. Opt.*, vol. 5, no. 7, pp. 1248–1249, 1966.

[118] C.-T. Li, S. Yin, and F. T. S. Yu, "Nonzero-order joint transform correlator," *Optical Engineering*, vol. 37, no. 1, pp. 58–65, 1998.

[119] Bahram Javidi, "Nonlinear joint power spectrum based optical correlation," *Appl. Opt.*, vol. 28, no. 12, pp. 2358–2367, 1989.

[120] J. A. Nelder and R. W. M. Wedderburn, "Generalized Linear Models," *Journal of the American Statistical Association*, vol. 135, no. 3, pp. 370–384, 1972.

[121] A. Reynaud, S. Takerkart, G. S. Masson, and F. Chavane, "Linear model decomposition for voltage-sensitive dye imaging signals : application in awake behaving monkey," *NeuroImage*, vol. 54, no. 2, pp. 1196–1210, 2010.

[122] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, vol. 1, pp. 511–518, 2001.

[123] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *International Conference on Image Processing*, vol. 1, pp. 900–903, 2002.

[124] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *Computer Vision–ECCV 2010*, pp. 749–762, Springer, 2010.

[125] P. Katz, M. Aron, and A. Alfalou, "Joint transform correlation for face tracking : elderly fall detection application," in *SPIE Defense, Security, and Sensing*, pp. 87480I–87480I, International Society for Optics and Photonics, 2013.

[126] P. Katz, M. Aron, and A. Alfalou, "A face-tracking system to detect falls in the elderly," *SPIE NEwsroom*, 2013.

[127] G. Bradski, "The opencv library," *Dr. Dobb's Journal of Software Tools*, vol. 25, no. 11, pp. 120–126, 2000.

[128] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition : A literature survey," *Acm Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[129] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[130] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer vision-eccv 2004*, pp. 469–481, Springer, 2004.

[131] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[132] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.

[133] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces : Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 19, no. 7, pp. 711–720, 1997.